

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2003, 13:02:49 ; Search time 104 Seconds  
(without alignments)  
6548.598 Million cell updates/sec

Title: US-09-891-138A-1  
Perfect score: 1543  
Sequence: 1 gctcctggcagagttttctg.....tgccataaataaatcaatata 1543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

		%								
Result		Query								
No.	Score	Match	Length	DB	ID	Description				
1	589.2	38.2	1996	2	US-08-559-524A-1	Sequence 1, Appli				
2	589.2	38.2	1996	3	US-08-749-707-1	Sequence 1, Appli				
3	88.4	5.7	1429	4	US-09-016-434-1068	Sequence 1068, Ap				
4	86.4	5.6	3055	4	US-09-016-434-1456	Sequence 1456, Ap				
5	85.4	5.5	2025	4	US-09-016-434-1482	Sequence 1482, Ap				
6	82.8	5.4	1571	4	US-09-016-434-1108	Sequence 1108, Ap				
7	82.2	5.3	1805	4	US-08-405-271A-18	Sequence 18, Appl				
8	82.2	5.3	1973	4	US-09-016-434-1391	Sequence 1391, Ap				
9	80	5.2	1586	1	US-08-461-244-1	Sequence 1, Appli				
10	80	5.2	1953	4	US-09-016-434-1096	Sequence 1096, Ap				
11	79.6	5.2	984	3	US-08-513-974B-57	Sequence 57, Appl				

12	79.6	5.2	984	4	US-09-461-436B-57	Sequence 57, Appl
13	79.6	5.2	1023	3	US-08-513-974B-379	Sequence 379, App
14	79.2	5.1	998	4	US-08-432-174A-3	Sequence 3, Appli
15	78.4	5.1	1495	4	US-09-016-434-1190	Sequence 1190, Ap
16	78.4	5.1	2156	1	US-08-012-988A-1	Sequence 1, Appli
17	78	5.1	1640	3	US-08-781-250-1	Sequence 1, Appli
18	77.6	5.0	1773	4	US-09-016-434-1405	Sequence 1405, Ap
19	76	4.9	1679	4	US-09-016-434-1097	Sequence 1097, Ap
20	76	4.9	2100	4	US-09-495-050A-289	Sequence 289, App
21	74.8	4.8	984	3	US-08-459-046-1	Sequence 1, Appli
22	74.8	4.8	984	4	US-09-102-710B-1	Sequence 1, Appli
23	74.2	4.8	1140	4	US-09-016-434-750	Sequence 750, App
24	74.2	4.8	1301	2	US-08-467-948A-7	Sequence 7, Appli
25	74.2	4.8	1301	3	US-08-467-947A-7	Sequence 7, Appli
26	74	4.8	1086	4	US-09-170-496D-77	Sequence 77, Appl
27	74	4.8	1086	4	US-09-170-496D-205	Sequence 205, App
28	74	4.8	1643	1	US-08-383-750-3	Sequence 3, Appli
29	74	4.8	1643	3	US-08-352-678-3	Sequence 3, Appli
30	74	4.8	1643	4	US-09-536-954-3	Sequence 3, Appli
31	74	4.8	1643	4	US-09-016-434-1279	Sequence 1279, Ap
32	74	4.8	1643	5	PCT-US93-09636-3	Sequence 3, Appli
33	73.8	4.8	1065	3	US-08-847-296B-2	Sequence 2, Appli
34	73.8	4.8	1071	4	US-08-567-882-6	Sequence 6, Appli
35	73.8	4.8	1116	4	US-08-720-565-5	Sequence 5, Appli
36	73.8	4.8	1193	4	US-08-720-565-3	Sequence 3, Appli
37	73.8	4.8	1201	4	US-09-016-434-1085	Sequence 1085, Ap
38	73.8	4.8	1689	4	US-08-720-565-1	Sequence 1, Appli
39	73.8	4.8	1915	3	US-08-575-967A-3	Sequence 3, Appli
40	72	4.7	1147	1	US-08-417-103-15	Sequence 15, Appl
41	72	4.7	1351	1	US-07-816-283-5	Sequence 5, Appli
42	72	4.7	1351	1	US-08-417-103-5	Sequence 5, Appli
43	72	4.7	1351	4	US-09-016-434-1303	Sequence 1303, Ap
44	71.4	4.6	1796	1	US-07-816-283-11	Sequence 11, Appl
45	71.4	4.6	1796	1	US-08-417-103-11	Sequence 11, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-559-524A-1

; Sequence 1, Application US/08559524A

; Patent No. 5871963

; GENERAL INFORMATION:

; APPLICANT: Conley, Pamela B.

; APPLICANT: Jantzen, Hans-Michael

; TITLE OF INVENTION: NOVEL PURINERGIC RECEPTOR

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN, LEWIS & BOCKIUS LLP

; STREET: 1800 M Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20036-5869

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/559,524A
;      FILING DATE:  15-NOV-1995
;      CLASSIFICATION:  435
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler, Reid G.
;      REGISTRATION NUMBER:  30,988
;      REFERENCE/DOCKET NUMBER:  044481-5010-00-US
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-467-7000
;      TELEFAX:  202-467-7176
;      INFORMATION FOR SEQ ID NO:  1:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  1996 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  cDNA
;      FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  625..1626
US-08-559-524A-1

```

```

Query Match          38.2%;  Score 589.2;  DB 2;  Length 1996;
Best Local Similarity 75.1%;  Pred. No. 5.3e-156;
Matches 762;  Conservative 0;  Mismatches 248;  Indels 4;  Gaps 2;

```

```

Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      632 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 691

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      692 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGAAATACCA 751

Qy      159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      752 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 811

Qy      219 TTAACCTTTCCATCTCTGACTTTGCTTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      812 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 871

Qy      279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      872 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 931

Qy      339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      932 ATGCCAACCTCTATACCAGCATTCCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 991

Qy      399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      992 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 1051

```





```

; ZIP: 20036-5869
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,707
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler, Reid G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 044481-5010-01-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-467-7000
; TELEFAX: 202-467-7176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1996 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 625..1626
US-08-749-707-1

```

```

Query Match          38.2%; Score 589.2; DB 3; Length 1996;
Best Local Similarity 75.1%; Pred. No. 5.3e-156;
Matches 762; Conservative 0; Mismatches 248; Indels 4; Gaps 2;

```

```

Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      632 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 691

Qy      99 AGTACTACCTCTCTGCATTTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      692 AGTACTACCTTTCCATTTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 751

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     752 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 811

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     812 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 871

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     872 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 931

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     932 ATGCCAACCTCTATACCAGCATTCTCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 991

```



```

; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1068:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1429 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1124904
US-09-016-434-1068

```

```

Query Match          5.7%; Score 88.4; DB 4; Length 1429;
Best Local Similarity 45.7%; Pred. No. 5.6e-15;
Matches 385; Conservative 0; Mismatches 451; Indels 6; Gaps 2;

```

```

Qy      107 CTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCACTGTGGTG 166
      || ||| ||||| | | || | | || || || || |
Db      292 CTGCCTGTGAGCTATGCAGTTGTCTTTGTGCTGGGCTTGGGCCTTAACGCCCCAACCTA 351

Qy      167 TTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTT 226
      | || | |||| | | || | | || | || || ||
Db      352 TGGCTCTTCATCTTCCGCCTCCGACCCTGGGATGCAACGGCCACCTACATGTTCCACCTG 411

Qy      227 TCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTAT---GCC 283
      | | || || | | || | || || || || || || || ||
Db      412 GCATTGTGAGACACCTTGTATGTGCTGTGCGCTGCCCCACCCTCATCTACTATTATGCAGCC 471

Qy      284 AATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACC 343
      | | | || || || || || || || | || | || | |
Db      472 CACAACCACTGGCCCTTTGGCACTGAGATCTGCAAGTTCGTCCGCTTTCTTTTCTATTGG 531

Qy      344 AACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATG 403

```

```

      ||||| ||| ||| ||||| | ||| ||| || ||| ||| |||
Db      532 AACCTCTACTGCAGTGTCTTTTCCTCACCTGCATCAGCGTGCACCGCTACCTGGGCATC 591

Qy      404 AAGTACCCTTTCCGAGAACACTTTCACAAAAGAAGGAATTTGCCATTTTAATCTCGCTG 463
      |||| | || | || | || | || | || | ||| |||
Db      592 TGCCACCCACTTCGGGCACCTACGCTGGGGCCGCCCTCGCCTCGCAGGCCTTCTCTGCCTG 651

Qy      464 GCTGTCTGGGCCTTAGTGACCTTAGAAGTTCACCCATGCTCACTTTCATCAATTCTGTC 523
      || || || | || || | | ||| || | || | || |
Db      652 GCAGTTTGGTTGGTCGTAGCCGGCTGCCTCGTGCCCAACCTGTTCTTTGTCAACAACAGC 711

Qy      524 CAAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCGGAAACCTGAACACAAT 583
      |||| | || ||| || | | ||| || | ||| ||| ||
Db      712 AACAAAGGGACCACCGTCTGTGCCATGACACCACTCGGCCTGAAGAGTTTGACCACTAT 771

Qy      584 CTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGC 643
      | | |||| | | || ||| | || | || |
Db      772 GTGCACTTCAGCTCGGCGGTTCATGGGGCTGCTCTTTGGCGTGCCCTGCCTGGTCACTCTT 831

Qy      644 TTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGCCAGCAGCAAGCAACTGCC 703
      | | ||| |||| | | | | | ||| |
Db      832 GTTTGCTATGGACTCATGGCTCGTCGCCTGTATCAGCCCTTGCCAGGCTCTGCACAGTCG 891

Qy      704 CTGCCACTGGACAAACCCCAACGCCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTC 763
      | | | | | || ||| || | ||| || | ||| || |
Db      892 TCTTCTCGCCTCCGCTCTCTCCGCACCATAGCTGTGGTGCTGACTGTCTTTGCTGTCTGC 951

Qy      764 TTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCA 823
      ||| || | || |||| |||| | | | ||| |
Db      952 TTCGTGCCTTTCCACATCACCCGCACCATTTACTACCTGGCCAGGCTGTTGGAA---GCT 1008

Qy      824 CAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTT 883
      | | | |||| |||| | || | |||| |||| |||| |
Db      1009 GACTGCCGAGTACTGAACATTGTCAACGTGGTCTATAAAGTGACTCGGCCCTTGCCAGT 1068

Qy      884 CTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATG 943
      |||| | | |||| | ||||| ||| || || | || |
Db      1069 GCCAACAGCTGCCTGGATCCTGTGCTCTACTTGCTCACTGGGGACAAATATCGACGTCAG 1128

Qy      944 CT 945
      ||
Db      1129 CT 1130

```

RESULT 4

US-09-016-434-1456

; Sequence 1456, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

```

; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1456:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3055 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g798835
US-09-016-434-1456

```

```

Query Match          5.6%; Score 86.4; DB 4; Length 3055;
Best Local Similarity 46.1%; Pred. No. 3.1e-14;
Matches 402; Conservative 0; Mismatches 461; Indels 9; Gaps 3;

```

```

Qy      80 ACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTT 139
      || || | || | | ||||| | || | || | |||| |
Db      982 ACCAAGACGGGCTTCCAGTTTTACTACCTGCCGGCTGTCTACATCTTGGTATTCATCATC 1041

Qy      140 GGA CTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAAC 199
      || | || || || | | | | | |||| | ||||| |
Db      1042 GGCTTCCTGGGCAACAGCGTGGCCATCTGGATGTTTCGTCTTCCACATGAAGCCCTGGAGC 1101

Qy      200 AGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTT 259
      ||| | || || | || || | | | |||| | || || || ||
Db      1102 GGCATCTCCGTGTACATGTTCAATTTGGCTCTGGCCGACTTCTTGACGTGCTGACTCTG 1161

Qy      260 CCCATCCTGATAAAGAGTTATGCCAATGATA---AGGGGACCTATGGAGATGTTCTCTGT 316
      || ||||| || |||| | | | ||| || || |||| | |||
Db      1162 CCAGCCCTGATCTTCTACTACTTCAATAAAACAGACTGGATCTTCGGGGATGCCATGTGT 1221

Qy      317 ATAAGCAACCGATATGTGCTTACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTC 376
      | | | | | |||| | ||||| | |||| | || || || |

```

Db 1222 AAACGAGAGGTTTCATCTTTCATGTGAACCTCTA---TGGCATCTTGTTTCTGACATGC 1278  
 Qy 377 ATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAG 436  
 || || |||| || || |||| || || || ||  
 Db 1279 ATCAGTGCCACCGGTACAGCGGTGTGGTGTACCCCTCAAGTCCCTGGGCCGGCTCAAA 1338  
 Qy 437 AAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTA 496  
 ||| | ||| || | ||| || |||| | ||| | | |  
 Db 1339 AAGAAGAATGCGATCTGTATCAGCGTGTGGTGTGGCTCATTGTGGTGGTGGCGATCTCC 1398  
 Qy 497 CCCATGCTCACTTTTCATCAATTCT---GTCCCAAAAGAAGAGGGCAGTAAGTGCATCGAC 553  
 |||| ||| | | | |||| ||| | | ||| ||||  
 Db 1399 CCCATCCTCTTCTACTCAGGTACCGGGGTCCGCAAAAACAAAACCATCACCTGTTACGAC 1458  
 Qy 554 TATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTG 613  
 | | ||| | |||| |||| | ||| ||| ||  
 Db 1459 ACCACCTCAGACGAGTACCTGCGAAGTTATTTTCATCTACAGCATGTGCACGACCGTGGCC 1518  
 Qy 614 GGCTTCTTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTA 673  
 ||| | || | |||| || || |||| || || ||  
 Db 1519 ATGTTCTGTGTCCCTTGGTGTGATTCTGGGCTGTTACGGATTAATTGTGAGAGCTTTG 1578  
 Qy 674 AAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACCCCAACGCCTGGTG 733  
 | | | | | | | | | | | | | | | | |  
 Db 1579 ATTTACAAAGATCTGGACAACCTCTCCTCTGAGGAGAAAATCGATTTACCTGGTAATCATT 1638  
 Qy 734 GTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTG 793  
 || ||| | ||| | |||| || | | | ||||  
 Db 1639 GTACTGACTGTTTTTGTGTGTCTTACATCCCTTTCCATGTGATGAAAACGATGAACCTTG 1698  
 Qy 794 AGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCT 853  
 ||| | | | || | | | | | | | | |  
 Db 1699 AGGGCCCGGCTTGATTTTCAGACCCAGCAATGTGTGCTTTCAATGACAGGGTTTATGCC 1758  
 Qy 854 ATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTAC 913  
 | || |||| | ||| || |||| ||| ||||  
 Db 1759 ACGTATCAGGTGACAAGAGGTCAGCAAGTCTCAACAGTTGTGTGGACCCCATCTCTAT 1818  
 Qy 914 TTCCTCATGGGAGACCATTACAGAGAGATGCT 945  
 ||| | |||| || |||| || ||  
 Db 1819 TTCTTGGCGGGAGATACTTTCAGAAGGAGACT 1850

RESULT 5

US-09-016-434-1482

; Sequence 1482, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

```

; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1482:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2025 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g984506
US-09-016-434-1482

```

```

Query Match          5.5%; Score 85.4; DB 4; Length 2025;
Best Local Similarity 46.5%; Pred. No. 4.7e-14;
Matches 389; Conservative 0; Mismatches 436; Indels 12; Gaps 3;

```

```

Qy      91 CTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGG 150
      ||| || |  || ||| | || | | ||| ||| |
Db      335 CTTCAAGTACGTGCTGCTGCCTGTGTCCTACGGCGTGGTGTGCGTGCTTGGGCTGTGTCT 394

Qy      151 GAATGTCACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGT 210
      ||| | | | | | | | | | | | | | | | |
Db      395 GAACGCCGTGGCGCTCTACATCTTCTTGTGCCGCCCTCAAGACCTGGAATGCGTCCACCAC 454

Qy      211 CTATCTTTTAACTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGAT 270
      ||| | || |||| | | |||| | | | |||| || | ||| |
Db      455 ATATATGTTCCACCTGGCTGTGTCTGATGCACTGTATGCGGCCCTCCCTGCCGCTGCTGGT 514

Qy      271 AAAGAGTTATGCCAAT--GATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCG 327
      |  || |||  || | || || |  || |||| |  ||
Db      515 CTATTACTACGCCCCGCGCGACCACTGGCCCTTCAGCACGGTGCTCTGCAAGCTGGTGCG 574

Qy      328 ATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTCATTAGCATGGA 387
      |  | | | ||||| ||| ||||| ||||| ||| ||| || |
Db      575 CTCCTCTTCTACACCAACCTTTACTGCAGCATCCTCTTCCTCACCTGCATCAGCGTGCA 634

```

Qy 388 CCGATATCTGCTCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGC 447  
 ||| | ||| | | ||| | | | | | | |  
 Db 635 CCGGTGTCTGGGCGTCTTACGACCTCTGCGCTCCCTGCGCTGGGGCCGGGCCCGCTACGC 694  
 Qy 448 CATTTTAATCTCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCAC 507  
 | | | | | | | | | | | | | | |  
 Db 695 TCGCCGGGTGGCCGGGGCCGTGTGGGTGTTGGTGCTGGCCTGCCAGGCCCCCGTGCTCTA 754  
 Qy 508 TTTTCATCAATTCTGTCCCAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCTGG 567  
 || ||| | | | | | | | | | | | | | | |  
 Db 755 CTTTGTCAACCACAGCGCGCGGGGGCCGCGTAACCTGCCACGACACCTCGGCACCCGA 814  
 Qy 568 AAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCC 627  
 | | | | | | | | | | | | | | | | | |  
 Db 815 GCTCTTCAGCCGCTTCGTGGCCTACAGCTCAGTCATGCTGGGCCTGCTCTTCGCGGTGCC 874  
 Qy 628 TCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCA 687  
 | | ||| | | ||| | | ||| | | | | | | |  
 Db 875 CTTTGCCGTCACTCCTTGTCTGTTACGTGCTCATGGCTCGGCGACTGCTAAAGCCAGCCTA 934  
 Qy 688 GCAGCAAGCAACTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGT 747  
 | | | | | | | | | | | | | | | | | |  
 Db 935 CGGGACCTCGGGCGGCCTCCCTAGGGCCAAGCGCAAGTCCGTGCGCACCATCGCCGTGGT 994  
 Qy 748 G-----ATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGC 801  
 | | ||| | | ||| | | | | | | | | | |  
 Db 995 GCTGGCTGTCTTCGCCCTCTGCTTCCTGCCATTCCACGTACCCGACCCCTCTACTACTC 1054  
 Qy 802 CTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACAC 861  
 || ||| | | | | | | | | | | | | | | |  
 Db 1055 CTTCCGCTCGCTGG---ACCTCAGCTGCCACACCCTCAACGCCATCAACATGGCCTACAA 1111  
 Qy 862 ACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCT 918  
 | | ||| | | ||| | | ||| | | | | | | |  
 Db 1112 GGTTACCCGGCCGCTGGCCAGTGCTAACAGTTGCCTTGACCCCGTGCTCTACTTCCT 1168

RESULT 6

US-09-016-434-1108

; Sequence 1108, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk



```

;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/016,434
;      FILING DATE:  HERewith
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Zeller, Karen J.
;      REGISTRATION NUMBER:  37,071
;      REFERENCE/DOCKET NUMBER:  PA-0002 US
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;      INFORMATION FOR SEQ ID NO:  1108:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  1571 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  g1296659
US-09-016-434-1108

```

```

Query Match          5.4%;  Score 82.8;  DB 4;  Length 1571;
Best Local Similarity 46.2%;  Pred. No. 2.2e-13;
Matches 390;  Conservative 0;  Mismatches 442;  Indels 12;  Gaps 3;

```

```

Qy      89 ATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTT 148
      | ||| || |      || | | ||| | | |      ||| |||
Db      343 AACTTCAAGCAACTGCTGCTGCCACCTGTGTATTTCGGCGGTGCTGGCGGCTGGCCTGCCG 402

Qy      149 GGAATGTCACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAAT 208
      ||| || ||| | | | | | | | | | | | | | | |
Db      403 CTGAACATCTGTGTCATTACCCAGATCTGCACGTCCCGCCGGGCCCTGACCCGCACGGCC 462

Qy      209 GTCTATCTTTTAAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTG 268
      || ||      | ||||| | | ||||| | |      ||| ||| ||| ||
Db      463 GTGTACACCCTAAACCTTGCTCTGGCTGACCTGCTATATGCCTGCTCCCTGCCCTGCTC 522

Qy      269 ATAAAGAGTTATGCCAA---TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAAC 325
      || | | ||||| | ||| | || ||| || | | ||      |
Db      523 ATCTACAACCTATGCCCAAGGTGATCACTGGCCCTTTGGCGACTTCGCCTGCCGCTGGTC 582

Qy      326 CGATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTCATTAGCATG 385
      || | | | | ||||| || ||||| ||||| ||||| ||| ||| |
Db      583 CGCTTCCTCTTCTATGCCAACCTGCACGGCAGCATCCTCTTCCTCACCTGCATCAGCTTC 642

Qy      386 GACCGATATCTGCTCATGAAGTACCCTTT---CCGAGAACACTTTCTACAAAAGAAGGAA 442
      | || ||| ||| ||| ||| | || ||| |
Db      643 CAGCGCTACCTGGGCATCTGCCACCCGCTGGCCCCCTGGCACAAACGTGGGGGGCCGCCG 702

```

Qy 443 TTTGCCATTTTAAATCTCGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATG 502  
 |||| | | | | | | | | |||| | | ||||  
 Db 703 GCTGCCTGGCTAGTGTGTGTAGCCGTGTGGCTGGCCGTGACAACCCAGTGCCTGCCCCACA 762

Qy 503 CTCACCTTTCATCAATTCTGTCCCAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGT 562  
 || ||| | | | | | ||| | ||| |||  
 Db 763 GCCATCTTCGCTGCCACAGGCATCCAGCGTAACCGCACTGTCTGCTATGACCTCAGCCCG 822

Qy 563 TCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTA 622  
 ||| | ||| || | || || | || ||| | | |||||  
 Db 823 CCTGCCCTGGCCACCCACTATATGCCCTATGGCATGGCTCTCACTGTCATCGGCTTCCTG 882

Qy 623 ATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGG 682  
 | || | ||| || || |||| ||| | | | | |  
 Db 883 CTGCCCTTTGCTGCCCTGCTGGCCTGTCTACTGTCTCCTGGCCTGCCGCTGTGCCGCCAG 942

Qy 683 AGCCAGCAGCAAGCAACTG-----CCCTGCCACTGGACAAACCCCAACGCCTGGTGGTC 736  
 | | || ||| || | | ||| | ||| ||| ||  
 Db 943 GATGGCCCGGCAGAGCCTGTGGCCAGGAGCGGCGTGGCAAGGCGGCCCGCATGGCCGTG 1002

Qy 737 CTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGG 796  
 ||| || || ||| | || ||| || | ||| |||| |  
 Db 1003 GTGGTGGCTGCTGCCTTTGCCATCAGCTTCCTGCCTTTTCACATCACCAAGACAGCCTAC 1062

Qy 797 ATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATA 856  
 | || ||| | | | | | | | |||| | |  
 Db 1063 CTGGCAGTGCCTCGACGCCGGCGTCCCCTGCACTGTATTGGAGGCCCTTTCAGCGGCC 1122

Qy 857 TACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTC 916  
 |||| | || |||| | || | |||| | | | ||||| ||| || |  
 Db 1123 TACAAAGGCACGCGGCCGTTTGCCAGTGCCAACAGCGTGCTGGACCCCATCCTCTTCTAC 1182

Qy 917 CTCA 920  
 |||  
 Db 1183 TTCA 1186

# RESULT 7

US-08-405-271A-18

; Sequence 18, Application US/08405271A

; Patent No. 6432652

## ; GENERAL INFORMATION:

; APPLICANT: EVANS, CHRISTOPHER J.

; APPLICANT: KEITH, DUANE E.

; TITLE OF INVENTION: OPIOID RECEPTOR GENES

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 2000 PENNSYLVANIA AVENUE, NW, Suite 5500

; CITY: WASHINGTON

; STATE: DC

; COUNTRY: USA

; ZIP: 20006-1888

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/405,271A
;   FILING DATE: 14-MAR-1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MURASHIGE, KATE H.
;   REGISTRATION NUMBER: 29,959
;   REFERENCE/DOCKET NUMBER: 22000-20526.22
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 887-1500
;   TELEFAX: (202) 887-0763
;   TELEX: 90-4030 MRSNFOERSWSH
;   INFORMATION FOR SEQ ID NO: 18:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1805 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 10..1119
US-08-405-271A-18

```

```

Query Match          5.3%; Score 82.2; DB 4; Length 1805;
Best Local Similarity 44.5%; Pred. No. 3.5e-13;
Matches 379; Conservative 0; Mismatches 463; Indels 9; Gaps 1;

```

```

Qy      85 GGCTATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACT 144
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147 GCCCCTCGGGCTCAAGGTCACCATCGTGGGGCTCTACCTGGCCGTGTGTGTCGGAGGGCT 206

Qy      145 GCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAG 204
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      207 CCTGGGGAACCTGCCTTGTATGTACGTCATCCTCAGGCACACCAAATGAAGACAGCCAC 266

Qy      205 CAATGTCTATCTTTTTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCCTTCCCAT 264
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      267 CAATATTTACATCTTTAACCTGGCCCTGGCCGACACTCTGGTCTGCTGACGCTGCCCTT 326

Qy      265 CCTGATAAAGAGTTATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAA 324
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      327 CCAGGGCACGGACATCCTCCTGGGCTTCTGGCCGTTTGGGAATGCGCTGTGCAAGACAGT 386

Qy      325 CCGATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCAT 384
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      387 CATTGCCATTGACTACTACAACATGTTCCACCAGCACCTTCACCCTAAGTCCCATGAGTGT 446

Qy      385 GGACCGATATCTGCTCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATT 444
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      447 GGATCGCTATGTAGCCATCTGCCACCCCATCCGTGCCCTCGACGTCCGCACGTCCAGCAA 506

Qy      445 TGCCATTTTAATCTCGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATGCT 504
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      507 AGCCCAGGCTGTCAATGTGGCCATCTGGGCCCTGGCCTCTGTTGTGCGGTGTTCCCGTTGC 566

Qy      505 CACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTC 564

```

Db 567 CATCATGGGCTCGGCACAGGTCGAGGATGAAGAGATCGAGTGCCTGGTGGAGATCCCTAC 626  
 Qy 565 TGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAAT 624  
 Db 627 CCCTCAGGATTACTGGGGCCCCGGTGTTCGCCATCTGCATCTTCCTCTTCTCCTTCATCGT 686  
 Qy 625 TCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAG 684  
 Db 687 CCCCCTGCTCGTCATCTCTGTCTGCTACAGCCTCATGATCCGGCGGCTCCGTGGAGTCCG 746  
 Qy 685 CCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGT 744  
 Db 747 CCTGCTCTCGGGCTCCCAGAGAAGGACCGGAACCTGCGGCGCATCACTCGGCTGGTGCT 806  
 Qy 745 TGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTC 804  
 Db 807 GGTGGTAGTGGCTGTGTTTCGTGGGCTGCTGGACGCCTGTCCAGGTCTTCGTGCTGGCCCA 866  
 Qy 805 ACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACT 864  
 Db 867 AGGGCTGGGGGTTTCAGCCGAGCAGCGAGACTGCCGTGGCCATTCTGCGCTTCTGCAC--- 923  
 Qy 865 GACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGG 924  
 Db 924 -----GGCCCTGGGCTACGTCAACAGCTGCCTCAACCCCATCCTCTACGCCTTCTGGA 977  
 Qy 925 AGACCATTACA 935  
 Db 978 TGAGAACTTCA 988

RESULT 8

US-09-016-434-1391

; Sequence 1391, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/016,434

; FILING DATE: HERewith

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1391:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g471316
US-09-016-434-1391

```

```

Query Match          5.3%; Score 82.2; DB 4; Length 1973;
Best Local Similarity 44.5%; Pred. No. 3.7e-13;
Matches 379; Conservative 0; Mismatches 463; Indels 9; Gaps 1;

```

```

Qy      85 GGCTATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACT 144
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      315 GCCCCTCGGGCTCAAGGTCAACCATCGTGGGGCTCTACCTGGCCGTGTGTGTCGGAGGGCT 374

Qy      145 GCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAG 204
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      375 CCTGGGGAACCTGCCTTGTCATGTACGTCATCCTCAGGCACACCAAATGAAGACAGCCAC 434

Qy      205 CAATGTCTATCTTTTAAACCTTTCCATCTCTGACTTTGCTTTTCTGTGCACCCTTCCCAT 264
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      435 CAATATTTACATCTTTAAACCTGGCCCTGGCCGACACTCTGGTCCTGCTGACGCTGCCCTT 494

Qy      265 CCTGATAAAGAGTTATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAA 324
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      495 CCAGGGCACGGACATCCTCCTGGGCTTCTGGCCGTTTGGGAATGCGCTGTGCAAGACAGT 554

Qy      325 CCGATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTCATTAGCAT 384
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      555 CATTGCCATTGACTACTACAACATGTTTACCAGCACCTTCACCCTAAGTCCCATGAGTGT 614

Qy      385 GGACCGATATCTGCTCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATT 444
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      615 GGATCGCTATGTAGCCATCTGCCACCCATCCGTGCCCTCGACGTCCGCACGTCCAGCAA 674

Qy      445 TGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCT 504
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      675 AGCCCAGGCTGTCAATGTGGCCATCTGGGCCCTGGCCTCTGTTGTCGGTGTTCCTGGTGC 734

Qy      505 CACTTTCATCAATTCTGTCCCAAAAGAAGAGGGGAGTAACTGCATCGACTATGCAAGTTC 564
      | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 735 CATCATGGGCTCGGCACAGGTCGAGGATGAAGAGATCGAGTGCCTGGTGGAGATCCCTAC 794  
 Qy 565 TGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAAT 624  
 | | | | | | | | | | | | | | | | | |  
 Db 795 CCCTCAGGATTACTGGGGCCCCGGTGTTCGCCATCTGCATCTTCCTCTCTCCTTCATCGT 854  
 Qy 625 TCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAG 684  
 | | | | | | | | | | | | | | | | | |  
 Db 855 CCCCCTGCTCGTCATCTCTGTCTGCTACAGCCTCATGATCCGGCGGCTCCGTGGAGTCCG 914  
 Qy 685 CCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGT 744  
 | | | | | | | | | | | | | | | | | |  
 Db 915 CCTGCTCTCGGGCTCCCAGAGAAGGACCGGAACCTGCGGCGCATCACTCGGCTGGTGCT 974  
 Qy 745 TGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTC 804  
 | | | | | | | | | | | | | | | | | |  
 Db 975 GGTGGTAGTGGCTGTGTTTCGTGGGCTGCTGGACGCCTGTCCAGGTCTTCGTGCTGGCCCA 1034  
 Qy 805 ACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACT 864  
 | | | | | | | | | | | | | | | | | |  
 Db 1035 AGGGCTGGGGGTTTCAGCCGAGCAGCGAGACTGCCGTGGCCATTCTGCGCTTCTGCAC--- 1091  
 Qy 865 GACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGG 924  
 | | | | | | | | | | | | | | | | | |  
 Db 1092 -----GGCCCTGGGCTACGTCAACAGCTGCCTCAACCCCATCCTCTACGCCTTCCTGGA 1145  
 Qy 925 AGACCATTACA 935  
 | | | | |  
 Db 1146 TGAGAACTTCA 1156

RESULT 9

US-08-461-244-1

; Sequence 1, Application US/08461244

; Patent No. 5776729

; GENERAL INFORMATION:

; APPLICANT: Soppet, Daniel R.

; APPLICANT: Yi, Li

; APPLICANT: Ruben, Steven M.

; APPLICANT: Rosen, Craig A.

; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HGBER32

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,

; ADDRESSEE: STUART & OLSTEIN

; STREET: 6 Becker Farm Road

; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/461,244

```

; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-445
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 431..1495
US-08-461-244-1

```

```

Query Match          5.2%; Score 80; DB 1; Length 1586;
Best Local Similarity 47.3%; Pred. No. 1.4e-12;
Matches 276; Conservative 0; Mismatches 305; Indels 3; Gaps 1;

```

```

Qy      98 AAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTC 157
      |||  |||  |||  |||||  ||  |||  ||  ||  ||  ||  ||  ||
Db      533 AAGTTGCTCCTTGCTGTCTTTTATTGCCTCCTGTTTGTATTTCAGTCTTCTGGGAAACAGC 592

Qy      158 ACTGTGGTGTTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTT 217
      ||  ||  ||  ||  ||  |||||  |||  ||  ||  ||  ||  ||
Db      593 CTGGTCATCCTGGTCCTTGTGGTCTGCAAGAAGCTGAGGAGCATCACAGATGTATACCTC 652

Qy      218 TTAAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGT 277
      ||  |||||  ||  ||  |||||  ||  |||  ||  ||  ||  ||  ||
Db      653 TTGAACCTGGCCCTGTCTGACCTGCTTTTTGTCTTCTCCTTCCCCTTTTCAGACCTA---C 709

Qy      278 TATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTT 337
      |||  ||  ||  ||  ||  |||  |||  ||  ||  ||  ||  ||
Db      710 TATCTGCTGGACCAAGTGGGTGTTTGGGACTGTAATGTGCAAAGTGGTGTCTGGCCTTTTAT 769

Qy      338 CACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTG 397
      |||  ||  |||||  |||||  ||  |||  |||  |||  ||  |||  |||
Db      770 TACATTGGCTTCTACAGCAGCATGTTTTTCATCACCCCTCATGAGTGTGGACAGGTACCTG 829

Qy      398 CTCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATC 457
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      830 GCTGTTGTCCATGCCGTGTATGCCCTAAAGGTGAGGACGATCAGGATGGGCACAACGCTG 889

Qy      458 TCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAAT 517
      ||  |||||  ||  |||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      890 TGCCTGGCAGTATGGCTAACC GCCATTATGGCTACCATCCCATTGCTAGTGT'TTTACCAA 949

Qy      518 TCTGTCCCAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCTGGAAACCCTGAA 577
      ||  ||  ||||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      950 GTGGCCTCTGAAGATGGTGTCTACAGTGTATTTCATTTTACAATCAACAGACTTTGAAG 1009

```

Qy 578 CACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTG 637  
 || || | || | || || | || || | || | || |  
 Db 1010 TGGAAGATCTTCACCAACTTCAAATGAACATTTTAGGCTTGTTGATCCCATTCACCATC 1069

Qy 638 ATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAG 681  
 | ||| ||| || || | | |||| |  
 Db 1070 TTTATGTTCTGCTACATTAAATCCTGCACCAGCTGAAGAGGTG 1113

RESULT 10

US-09-016-434-1096

; Sequence 1096, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/016,434

; FILING DATE: HERewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Zeller, Karen J.

; REGISTRATION NUMBER: 37,071

; REFERENCE/DOCKET NUMBER: PA-0002 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166

; INFORMATION FOR SEQ ID NO: 1096:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1953 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GENBANK

; CLONE: g1245056

US-09-016-434-1096



Query Match 5.2%; Score 80; DB 4; Length 1953;  
Best Local Similarity 47.3%; Pred. No. 1.5e-12;  
Matches 276; Conservative 0; Mismatches 305; Indels 3; Gaps 1;

```
Qy      98 AAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTC 157
      ||| | | | | | | | | | | | | | | | | | | | | |
Db     369 AAGTTGCTCCTTGCTGTCTTTTATTGCCTCCTGTTTGTATTTCAGTCTTCTGGGAAACAGC 428

Qy     158 ACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTT 217
      || | | | | | | | | | | | | | | | | | | | |
Db     429 CTGGTCATCCTGGTCCTTGTGGTCTGCAAGAAGCTGAGGAGCATCACAGATGTATACCTC 488

Qy     218 TTTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGT 277
      || | | | | | | | | | | | | | | | | | | | |
Db     489 TTGAACCTGGCCCTGTCTGACCTGCTTTTTTGTCTTCTCCTTCCCCTTTCAGACCTA---C 545

Qy     278 TATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTT 337
      || | | | | | | | | | | | | | | | | | | | |
Db     546 TATCTGCTGGACAGTGGGTGTTTGGGACTGTAATGTGCAAAGTGGTGTCTGGCTTTTAT 605

Qy     338 CACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTATTAGCATGGACCGATATCTG 397
      || | | | | | | | | | | | | | | | | | | | |
Db     606 TACATTGGCTTCTACAGCAGCATGTTTTTCATCACCTCATGAGTGTGGACAGGTACCTG 665

Qy     398 CTCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATC 457
      | | | | | | | | | | | | | | | | | | | | |
Db     666 GCTGTTGTCCATGCCGTGTATGCCCTAAAGGTGAGGACGATCAGGATGGGCACAACGCTG 725

Qy     458 TCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAAT 517
      | | | | | | | | | | | | | | | | | | | | |
Db     726 TGCCTGGCAGTATGGCTAACGCCATTATGGCTACCATCCCATTGCTAGTGTTTTACCAA 785

Qy     518 TCTGTCCCAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAA 577
      | | | | | | | | | | | | | | | | | | | | |
Db     786 GTGGCCTCTGAAGATGGTGTCTACAGTGTATTTCATTTTACAATCAACAGACTTTGAAG 845

Qy     578 CACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTG 637
      || | | | | | | | | | | | | | | | | | | | |
Db     846 TGGAAGATCTTCACCAACTTCAAATGAACATTTTAGGCTTGTGATCCCATTCCACCATC 905

Qy     638 ATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAG 681
      | | | | | | | | | | | | | | | | | | | | |
Db     906 TTTATGTTCTGCTACATTAAATCCTGCACCAGCTGAAGAGGTG 949
```

RESULT 11

US-08-513-974B-57

; Sequence 57, Application US/08513974B

; Patent No. 6114139

; GENERAL INFORMATION:

; APPLICANT: Hinuma, Shuji

; APPLICANT: Hosoya, Masaki

; APPLICANT: Fujii, Ryo

; APPLICANT: Ohtaki, Tetsuya

; APPLICANT: Fukusumi, Shoji

; APPLICANT: Ohgi, Kazuhiro

; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,

; TITLE OF INVENTION: PRODUCTION, AND USE THEREOF  
; NUMBER OF SEQUENCES: 380  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/513,974B  
; FILING DATE: 14-SEP-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP95/01599  
; FILING DATE: 10-AUG-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-093989  
; FILING DATE: 19-AUG-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-057186  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-007177  
; FILING DATE: 20-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-326611  
; FILING DATE: 28-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-270017  
; FILING DATE: 02-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-236357  
; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-236356  
; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-189274  
; FILING DATE: 11-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-189273  
; FILING DATE: 11-AUG-1945  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-189272  
; FILING DATE: 11-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Resnick, David S.  
; REGISTRATION NUMBER: 34,235  
; REFERENCE/DOCKET NUMBER: 45753  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400

```
; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-513-974B-57
```

```
Query Match          5.2%; Score 79.6; DB 3; Length 984;
Best Local Similarity 46.0%; Pred. No. 1.4e-12;
Matches 388; Conservative 0; Mismatches 444; Indels 12; Gaps 3;
```

```
Qy      89 ATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTT 148
      | | | | | | | | | | | | | | | | | | | | | |
Db      67 AACTTCAAGCAACTGCTGCTGCCACCTGTGTATTTCGGCGGTGCTGGCGGCTGGCCTGCCG 126

Qy     149 GGAATGTCACCTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAAT 208
      | | | | | | | | | | | | | | | | | | | | | |
Db     127 CTGAACATCTGTGTTCATTACCCAGATCTGCACGTCCCGCCGGGCCCTGACCCGCACGGCC 186

Qy     209 GTCTATCTTTTTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTG 268
      | | | | | | | | | | | | | | | | | | | | | |
Db     187 GTGTACACCCTAAACCTTGCTCTGGCTGACCTGCTATATGCCTGCTCCCTGCCCTGCTC 246

Qy     269 ATAAAGAGTTATGCCAA---TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAAC 325
      | | | | | | | | | | | | | | | | | | | | | |
Db     247 ATCTACAACATATGCCCAAGGTGATCACTGGCCCTTTGGCGACTTCGCCTGCCGCTGGTC 306

Qy     326 CGATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTATTAGCATG 385
      | | | | | | | | | | | | | | | | | | | | | |
Db     307 CGCTTCCTCTTCTATGCCAACCTGCACGGCAGCATCCTCTTCCTCACTGCATCAGCTTC 366

Qy     386 GACCGATATCTGCTCATGAAGTACCCTTT---CCGAGAACACTTTCTACAAAAGAAGGAA 442
      | | | | | | | | | | | | | | | | | | | | | |
Db     367 CAGCGCTACCTGGGCATCTGCCACCCGCTGGCCCCCTGGCACAAACGTGGGGGCCGCCGG 426

Qy     443 TTTGCCATTTTAAATCTCGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATG 502
      | | | | | | | | | | | | | | | | | | | | | |
Db     427 GCTGCCTGGCTAGTGTGTGTAACCGTGTGGCTGGCCGTGACAACCCAGTGCCTGCCCACA 486

Qy     503 CTCACCTTTCATCAATTCTGTCCCAAAAGAAGAGGGCAGTAACCTGCATCGACTATGCAAGT 562
      | | | | | | | | | | | | | | | | | | | | | |
Db     487 GCCATCTTCGCTGCCACAGGCATCCAGCGTAACCGCACTGTCTGCTATGACCTCAGCCCG 546

Qy     563 TCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTA 622
      | | | | | | | | | | | | | | | | | | | | | |
Db     547 CCTGCCCTGGCCACCCACTATATGCCCTATGGCATGGCTCTCACTGTCATCGGCTTCCTG 606

Qy     623 ATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGG 682
      | | | | | | | | | | | | | | | | | | | | | |
Db     607 CTGCCCTTTGCTGCCCTGCTGGCCTGCTACTGTCTCCTGGCCTGCCGCTGTGCCGCCAG 666

Qy     683 AGCCAGCAGCAAGCAACTG-----CCCTGCCACTGGACAAACCCCAACGCCTGGTGGTC 736
      | | | | | | | | | | | | | | | | | | | | | |
Db     667 GATGGCCCGGCAGAGCCTGTGGCCAGGAGCGGCGTGGCAAGGCGGCCGATGGCCGTG 726
```

Qy 737 CTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGG 796  
 ||| ||| || ||| | || |||| || | ||| |||| |  
 Db 727 GTGGTGGCTGCTGCCTTTGCCATCAGCTTCCTGCCTTTTCACATCACCAAGACAGCCTAC 786  
 Qy 797 ATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATA 856  
 | || | | | | | | | | | | | | | | | | |  
 Db 787 CTGGCAGTGGGCTCGACGCCGGGCGTCCCCTGCACTGTATTGGAGGCCTTTGCAGCGGCC 846  
 Qy 857 TACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTC 916  
 |||| | || |||| | || | |||| | | |||| || || |  
 Db 847 TACAAAGGCACGCGGCCGTTTGCCAGTGCCAACAGCGTGCTGGACCCCATCCTCTTCTAC 906  
 Qy 917 CTCA 920  
 |||  
 Db 907 TTCA 910

RESULT 12

US-09-461-436B-57

; Sequence 57, Application US/09461436B

; Patent No. 6538107

; GENERAL INFORMATION:

; APPLICANT: Shuji Hinuma

; Yasuaki Ito

; Ryo Fujii

; TITLE OF INVENTION: G Protein Coupled Receptor Protein,  
 ; Production, And Use Thereof

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Edwards & Angell, LLP

; STREET: 101 Federal Street

; CITY: BOSTON

; STATE: MA

; COUNTRY: USA

; ZIP: 02209

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/461,436B

; FILING DATE: 14-Dec-1999

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/513,974

; FILING DATE: 14-SEP-1995

; APPLICATION NUMBER: PCT/JP95/01599

; FILING DATE: 10-AUG-1995

; APPLICATION NUMBER: 7-093989

; FILING DATE: 19-APR-1995

; APPLICATION NUMBER: 7-057186

; FILING DATE: 16-MAR-1995

; APPLICATION NUMBER: 7-007177

; FILING DATE: 20-JAN-1995

; APPLICATION NUMBER: 6-326611

```

; FILING DATE: 28-DEC-1994
; APPLICATION NUMBER: 6-270017
; FILING DATE: 02-NOV-1994
; APPLICATION NUMBER: 6-236357
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: 6-236356
; FILING DATE: 30-SEP-1994
; APPLICATION NUMBER: 6-189274
; FILING DATE: 11-AUG-1994
; APPLICATION NUMBER: 6-189273
; FILING DATE: 11-AUG-1994
; APPLICATION NUMBER: 6-189272
; FILING DATE: 11-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: CONLIN, DAVID G.
; REGISTRATION NUMBER: <Unknown>
; REFERENCE/DOCKET NUMBER: 45753 DIV2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-439-4444
; TELEFAX: 617-439-4170
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-09-461-436B-57

```

```

Query Match          5.2%; Score 79.6; DB 4; Length 984;
Best Local Similarity 46.0%; Pred. No. 1.4e-12;
Matches 388; Conservative 0; Mismatches 444; Indels 12; Gaps 3;

```

```

Qy      89 ATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTT 148
      | ||| || | | | | | | | | | | | | | | | |
Db      67 AACTTCAAGCAACTGCTGCTGCCACCTGTGTATTGCGCGGTGCTGGCGGTGGCCTGCCG 126

Qy      149 GGAATGTCACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAAT 208
      ||| || ||| | | | | | | | | | | | | | |
Db      127 CTGAACATCTGTGTCACTACCCAGATCTGCACGTCCCGCCGGGCCCTGACCCGCACGGCC 186

Qy      209 GTCTATCTTTTAAACCTTTCCATCTCTGACTTTGCTTTTCCTGTGCACCCTTCCCATCCTG 268
      || || | ||||| | | |||| | | | ||| ||| ||| ||
Db      187 GTGTACACCCTAAACCTTGCTCTGGCTGACCTGCTATATGCCTGCTCCCTGCCCTGCTC 246

Qy      269 ATAAAGAGTTATGCCAA---TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAAC 325
      || | | ||||| | |||| | || ||| ||| || | ||
Db      247 ATCTACAACATATGCCCAAGGTGATCACTGGCCCTTTGGCGACTTCGCTGCCGCTGGTC 306

Qy      326 CGATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTCATTAGCATG 385
      || | | | | ||||| || ||||| ||||| ||||| ||| |||
Db      307 CGCTTCCTCTTCTATGCCAACCTGCACGGCAGCATCCTCTTCCTCACCTGCATCAGCTTC 366

Qy      386 GACCGATATCTGCTCATGAAGTACCCTTT---CCGAGAACACTTTCTACAAAAGAAGGAA 442
      | || || ||| ||| |||| | || ||| |
Db      367 CAGCGCTACCTGGGCATCTGCCACCCGCTGGCCCCCTGGCACAAACGTGGGGGCCGCCGG 426

```

Qy 443 TTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATG 502  
 |||| | | | | | | | | | | | | | |  
 Db 427 GCTGCCTGGCTAGTGTGTGTAACCGTGTGGCTGGCCGTGACAACCCAGTGCCTGCCACA 486  
 Qy 503 CTCAC'TTTCATCAATTCTGTCCCAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGT 562  
 || ||| | | | | | | | | | | |  
 Db 487 GCCATCTTCGCTGCCACAGGCATCCAGCGTAACCGCACTGTCTGCTATGACCTCAGCCCG 546  
 Qy 563 TCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTA 622  
 ||| | | | | | | | | | | | | | |  
 Db 547 CCTGCCCTGGCCACCCACTATATGCCCTATGGCATGGCTCTCACTGTCATCGGCTTCCTG 606  
 Qy 623 ATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGG 682  
 | | | | | | | | | | | | | | | | |  
 Db 607 CTGCCCTTTGCTGCCCTGCTGGCCTGCTACTGTCTCCTGGCCTGCCGCCTGTGCCGCCAG 666  
 Qy 683 AGCCAGCAGCAAGCAACTG-----CCCTGCCACTGGACAAACCCCAACGCCTGGTGGTC 736  
 | | | | | | | | | | | | | | | |  
 Db 667 GATGGCCCGGCAGAGCCTGTGGCCAGGAGCGGCGTGGCAAGGCGGCCCGCATGGCCGTG 726  
 Qy 737 CTGGCGGT'TGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGG 796  
 ||| ||| | | | | | | | | | | | |  
 Db 727 GTGGTGGCTGCTGCCTTTGCCATCAGCTTCCTGCCTTTTCACATCACCAAGACAGCCTAC 786  
 Qy 797 ATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATA 856  
 | | | | | | | | | | | | | | | |  
 Db 787 CTGGCAGTGGGCTCGACGCCGGGCGTCCCTGCACTGTATTGGAGGCCTTTGCAGCGGCC 846  
 Qy 857 TACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTC 916  
 |||| | | | | | | | | | | | | | |  
 Db 847 TACAAAGGCACGCGGCCGTTTGCAGTGCCAAACAGCGTGTGGACCCCATCCTCTTCTAC 906  
 Qy 917 CTCA 920  
 |||  
 Db 907 TTCA 910

RESULT 13

US-08-513-974B-379

; Sequence 379, Application US/08513974B

; Patent No. 6114139

; GENERAL INFORMATION:

; APPLICANT: Hinuma, Shuji

; APPLICANT: Hosoya, Masaki

; APPLICANT: Fujii, Ryo

; APPLICANT: Ohtaki, Tetsuya

; APPLICANT: Fukusumi, Shoji

; APPLICANT: Ohgi, Kazuhiro

; TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,

; TITLE OF INVENTION: PRODUCTION, AND USE THEREOF

; NUMBER OF SEQUENCES: 380

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

; STREET: 130 Water Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA  
;  
; ZIP: 02109  
;  
; COMPUTER READABLE FORM:  
;  
; MEDIUM TYPE: Floppy disk  
;  
; COMPUTER: IBM PC compatible  
;  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
;  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;  
; CURRENT APPLICATION DATA:  
;  
; APPLICATION NUMBER: US/08/513,974B  
;  
; FILING DATE: 14-SEP-1995  
;  
; CLASSIFICATION: 536  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: PCT/JP95/01599  
;  
; FILING DATE: 10-AUG-1995  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 7-093989  
;  
; FILING DATE: 19-AUG-1995  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 7-057186  
;  
; FILING DATE: 16-MAR-1995  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 7-007177  
;  
; FILING DATE: 20-JAN-1995  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-326611  
;  
; FILING DATE: 28-DEC-1994  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-270017  
;  
; FILING DATE: 02-NOV-1994  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-236357  
;  
; FILING DATE: 30-SEP-1994  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-236356  
;  
; FILING DATE: 30-SEP-1994  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-189274  
;  
; FILING DATE: 11-AUG-1994  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-189273  
;  
; FILING DATE: 11-AUG-1945  
;  
; PRIOR APPLICATION DATA:  
;  
; APPLICATION NUMBER: JP 6-189272  
;  
; FILING DATE: 11-AUG-1994  
;  
; ATTORNEY/AGENT INFORMATION:  
;  
; NAME: Resnick, David S.  
;  
; REGISTRATION NUMBER: 34,235  
;  
; REFERENCE/DOCKET NUMBER: 45753  
;  
; TELECOMMUNICATION INFORMATION:  
;  
; TELEPHONE: 617-523-3400  
;  
; TELEFAX: 617-523-6440  
;  
; INFORMATION FOR SEQ ID NO: 379:  
;  
; SEQUENCE CHARACTERISTICS:  
;  
; LENGTH: 1023 base pairs  
;  
; TYPE: nucleic acid  
;  
; STRANDEDNESS: double  
;  
; TOPOLOGY: linear

; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 37..1020  
US-08-513-974B-379

Query Match 5.2%; Score 79.6; DB 3; Length 1023;  
Best Local Similarity 46.0%; Pred. No. 1.4e-12;  
Matches 388; Conservative 0; Mismatches 444; Indels 12; Gaps 3;

```
Qy      89 ATCTTGAATAAGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTT 148
      | ||| || |          || | | ||| | | | | ||| |||
Db      103 AACTTCAAGCAACTGCTGCTGCCACCTGTGTATTCGGCGGTGCTGGCGGCTGGCCTGCCG 162

Qy      149 GGAATGTCACTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAAT 208
      ||| || ||| | | | | | | | | | | | | ||| |||
Db      163 CTGAACATCTGTGTCAATTACCCAGATCTGCACGTCCCGCCGGGCCCTGACCCGCACGGCC 222

Qy      209 GTCTATCTTTTTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTG 268
      || || | ||||| | | ||||| | | ||| ||| ||| |||
Db      223 GTGTACACCCTAAACCTTGCTCTGGCTGACCTGCTATATGCCTGCTCCCTGCCCTGCTC 282

Qy      269 ATAAAGAGTTATGCCAA---TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAAC 325
      || | | ||||| | |||| | |||| ||| || | |||
Db      283 ATCTACAACTATGCCCAAGGTGATCACTGGCCCTTTGGCGACTTCGCCTGCCGCCTGGTC 342

Qy      326 CGATATGTGCTTCACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTCATTAGCATG 385
      || | | | | ||||| || ||||| ||||| ||| ||| |
Db      343 CGCTTCTCTTCTATGCCAACCTGCACGGCAGCATCCTCTTCCTCACCTGCATCAGCTTC 402

Qy      386 GACCGATATCTGCTCATGAAGTACCCTTT---CCGAGAACACTTTCTACAAAAGAAGGAA 442
      | ||| ||| ||| |||| | || ||| |
Db      403 CAGCGCTACCTGGGCATCTGCCACCCGCTGGCCCCCTGGCACAAACGTGGGGGCCGCCGG 462

Qy      443 TTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATG 502
      |||| ||| | | | |||| |||| | || ||||
Db      463 GCTGCCTGGCTAGTGTGTGTAACCGTGTGGCTGGCCGTGACAACCCAGTGCCTGCCACA 522

Qy      503 CTCACCTTTCATCAATTCTGTCCCAAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGT 562
      || ||| | | | | | ||| | ||| |||
Db      523 GCCATCTTCGCTGCCACAGGCATCCAGCGTAACCGCACTGTCTGCTATGACCTCAGCCCCG 582

Qy      563 TCTGGAAACCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTA 622
      ||| | | ||| || | || || | ||| | | |||||
Db      583 CCTGCCCTGGCCACCCACTATATGCCCTATGGCATGGCTCTCACTGTCATCGGCTTCCTG 642

Qy      623 ATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGG 682
      | || | ||| || || |||| | || | | |
Db      643 CTGCCCTTTGCTGCCCTGCTGGCCTGCTACTGTCTCCTGGCCTGCCGCCTGTGCCGCCAG 702

Qy      683 AGCCAGCAGCAAGCAACTG-----CCCTGCCACTGGACAAACCCCAACGCCTGGTGGTC 736
      | | || ||| || | | ||| | ||| ||| ||
Db      703 GATGGCCCGGCAGAGCCTGTGGCCCAGGAGCGGCGTGGCAAGGCGGCCCGCATGGCCGTG 762

Qy      737 CTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGG 796
      ||| || || ||| | || |||| || | ||| ||| |
Db      763 GTGGTGGCTGCTGCCTTTGCCATCAGCTTCCTGCCTTTTCACATCACCAAGACAGCCTAC 822
```



```

Qy      797 ATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATA 856
      | | | | | | | | | | | | | | | | | |
Db      823 CTGGCAGTGGGCTCGACGCCGGGCGTCCCCTGCACTGTATTGGAGGCCTTTGCAGCGGCC 882

Qy      857 TACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTC 916
      | | | | | | | | | | | | | | | | | |
Db      883 TACAAAGGCACGCGGCCGTTTGCCAGTGCCAAACAGCGTGCTGGACCCCATCCTCTTCTAC 942

Qy      917 CTCA 920
      | | |
Db      943 TTCA 946

```

# RESULT 14

US-08-432-174A-3

```

; Sequence 3, Application US/08432174A
; Patent No. 6562587
; GENERAL INFORMATION:
; APPLICANT: KIEFFER, BRIGITTE
; TITLE OF INVENTION: NOVEL POLYPEPTIDES HAVING OPIOID RECEPTOR ACTIVITY,
; TITLE OF INVENTION: NUCLEIC ACIDS CODING THEREFOR AND USES THEREOF
; FILE REFERENCE: EX92009-US
; CURRENT APPLICATION NUMBER: US/08/432,174A
; CURRENT FILING DATE: 1995-05-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(996)
; NAME/KEY: modified_base
; LOCATION: (922)
; OTHER INFORMATION: a, t, c, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (927)
; OTHER INFORMATION: a, t, c, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (931)..(932)
; OTHER INFORMATION: a, t, c, g, other or unknown

```

US-08-432-174A-3

```

Query Match          5.1%; Score 79.2; DB 4; Length 998;
Best Local Similarity 50.0%; Pred. No. 1.8e-12;
Matches 198; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

```

Qy      106 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACGTGTGGT 165
      | | | | | | | | | | | | | | | | | |
Db      33 CATCACCGCGCTCTACTCGGCCGTGTGCGCCGTGGGGCTGCTGGGCAACGTGCTTGTTCAT 92

Qy      166 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCT 225
      | | | | | | | | | | | | | | | | | |
Db      93 GTTCGGCATCGTCCGGTACACTAAGATGAAGACGACCACCAAGATCTACATCTTCAACCT 152

```

Qy 226 TTCCATCTCTGACTTTGCTTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA 285  
 || | | || | ||| || || ||| || | || |  
 Db 153 GGCCTTAGCCGATGCGCTGGCCACCAGCACGCTGCCTTTCCAGAGTGCCAAGTACCTGAT 212  
 Qy 286 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA 345  
 || | | || || || || | |||| | | | || |||  
 Db 213 GGAGACGTGGCCCTTCGGCGAGCTGCTCTGCAAGGCTGTGCTCTCCATCGACTACTACAA 272  
 Qy 346 CCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAA 405  
 | | ||||| || || || | ||| || | ||| || |  
 Db 273 TATGTTACCAGCATCTTCACGCTCACCATGATGAGTGTGACCGCTACATCGCTGTCTG 332  
 Qy 406 GTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGC 465  
 |||| | | | | | | | | ||| | || |  
 Db 333 CCACCCTGTCAAGGCCCTGGACTTCCGCACGCCTGCCAAGGCCAAGCTGATCAACATCTG 392  
 Qy 466 TGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCAT 501  
 | |||| | | | | | | | |||||  
 Db 393 TATCTGGGTCTTGGCCTCAGGCGTTGGCGTGCCCAT 428

RESULT 15

US-09-016-434-1190

; Sequence 1190, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/016,434

; FILING DATE: HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Zeller, Karen J.

; REGISTRATION NUMBER: 37,071

; REFERENCE/DOCKET NUMBER: PA-0002 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

```
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1190:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1495 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g179984
US-09-016-434-1190
```

```
Query Match          5.1%; Score 78.4; DB 4; Length 1495;
Best Local Similarity 54.1%; Pred. No. 3.8e-12;
Matches 160; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
```

```
Qy      107 CTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCACCTGTGGTG 166
          || | | | || | | | || | ||| ||| ||| || | ||| |||
Db      109 CTGCCCCCTCTGTACTCCTTGGTATTTGTCAATTGGCCTGGTTGGAAACATCCTGGTGGTC 168

Qy      167 TTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTT 226
          || | | | | || | || | || | ||| || | ||| || | ||| |||
Db      169 CTGGTCCTTGTGCAATAACAAGAGGCTAAAAACATGACCAGCATCTACCTCCTGAACCTG 228

Qy      227 TCCATCTCTGACTTTTGCTTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT 286
          |||| ||||| | ||||| ||| ||||| || ||| | | ||| ||
Db      229 GCCATTTCTGACCTGCTCTTCCTGTTCACGCTTCCCTTCTGGATCGACTACAAGTTGAAG 288

Qy      287 GATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAAC 346
          ||| | || | ||| |||| | |||| | | | | | ||| || |
Db      289 GATGACTGGGTTTTTGGTGATGCCATGTGTAAGATCCTCTCTGGGTTTTATTACACAGGC 348

Qy      347 CTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCAT 402
          | |||| | ||| | ||| ||| | | | ||| ||| ||| ||| |||
Db      349 TTGTACAGCGAGATCTTTTTCATCATCCTGCTGACGATTGACAGGTACCTGGCCAT 404
```

```
Search completed: December 14, 2003, 15:02:12
Job time : 107 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2003, 15:00:25 ; Search time 516 Seconds  
(without alignments)  
9938.592 Million cell updates/sec

Title: US-09-891-138A-1  
Perfect score: 1543  
Sequence: 1 gctcctggcagagttttctg.....tgctaaataaatcaatata 1543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result	Query					
No.	Score	Match	Length	DB	ID	Description

1	1543	100.0	1543	11	US-09-891-138A-1	Sequence 1, Appli
2	592.4	38.4	1005	13	US-10-272-983-35	Sequence 35, Appl
3	592.4	38.4	1005	13	US-10-393-807-35	Sequence 35, Appl
4	592.4	38.4	1380	15	US-10-225-567A-566	Sequence 566, App
5	592.4	38.4	1436	13	US-09-764-886-36	Sequence 36, Appl
6	592.4	38.4	4232	13	US-09-764-886-11	Sequence 11, Appl
7	590.8	38.3	1428	15	US-10-270-587-1	Sequence 1, Appli
8	126.6	8.2	1014	9	US-09-943-798-3	Sequence 3, Appli
9	126.6	8.2	1014	11	US-09-885-453-2	Sequence 2, Appli
10	126.6	8.2	1014	13	US-10-321-807-27	Sequence 27, Appl
11	126.6	8.2	1014	15	US-10-270-144-1	Sequence 1, Appli
12	126.6	8.2	1014	15	US-10-188-405-7	Sequence 7, Appli
13	126.6	8.2	1014	15	US-10-079-384-13	Sequence 13, Appl
14	126.6	8.2	1014	15	US-10-225-567A-646	Sequence 646, App
15	126.6	8.2	1081	13	US-10-010-568-1	Sequence 1, Appli
16	126.6	8.2	1414	13	US-10-017-161-525	Sequence 525, App
17	126.6	8.2	9905	15	US-10-270-144-3	Sequence 3, Appli
18	125	8.1	1014	13	US-10-278-141-10	Sequence 10, Appl
19	125	8.1	1014	13	US-10-296-081-10	Sequence 10, Appl
20	125	8.1	1014	15	US-10-023-775B-1	Sequence 1, Appli
21	119.2	7.7	1020	11	US-09-782-974C-85	Sequence 85, Appl
22	119	7.7	1313	10	US-09-728-422-1	Sequence 1, Appli
23	104	6.7	1020	13	US-10-023-586B-3	Sequence 3, Appli
24	100.8	6.5	2424	15	US-10-225-567A-218	Sequence 218, App
25	97.8	6.3	1002	13	US-10-088-726-25	Sequence 25, Appl
26	97.8	6.3	1082	13	US-10-023-586B-1	Sequence 1, Appli
27	97.8	6.3	1334	13	US-10-017-161-707	Sequence 707, App
28	97.8	6.3	3143	13	US-10-275-910-1	Sequence 1, Appli
29	96	6.2	831	9	US-09-943-798-1	Sequence 1, Appli
30	95.4	6.2	1017	11	US-09-885-453-3	Sequence 3, Appli
31	95.4	6.2	1017	15	US-10-079-384-5	Sequence 5, Appli
32	90	5.8	1651	13	US-10-366-288-41	Sequence 41, Appl
33	88.4	5.7	1098	15	US-10-225-567A-331	Sequence 331, App
34	88.4	5.7	1429	11	US-09-077-173A-1	Sequence 1, Appli
35	87	5.6	2118	13	US-10-101-510-431	Sequence 431, App
36	85.4	5.5	2025	13	US-10-101-510-722	Sequence 722, App
37	85.4	5.5	2025	15	US-10-225-567A-216	Sequence 216, App
38	84.8	5.5	2796	13	US-10-024-298A-173	Sequence 173, App
39	84.8	5.5	2796	13	US-10-042-211A-173	Sequence 173, App
40	83.8	5.4	1113	13	US-09-826-509-540	Sequence 540, App
41	83	5.4	850	13	US-10-275-910-4	Sequence 4, Appli
42	83	5.4	851	13	US-10-275-910-6	Sequence 6, Appli
43	82.8	5.4	1571	13	US-10-354-358-9	Sequence 9, Appli
44	82.8	5.4	1832	15	US-10-225-567A-222	Sequence 222, App
45	82.2	5.3	1805	9	US-09-823-114-18	Sequence 18, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-891-138A-1

; Sequence 1, Application US/09891138A

; Publication No. US20030083245A1

; GENERAL INFORMATION:

; APPLICANT: Lin, Daniel Chi-Hong

```
; APPLICANT: Zhao, Jiagang
; APPLICANT: Chen, Jin-Long
; APPLICANT: Cutler, Gene
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: No. US20030083245A1el Receptors
; FILE REFERENCE: 018781-006210US
; CURRENT APPLICATION NUMBER: US/09/891,138A
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,461
; PRIOR FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1543
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (44)..(997)
; OTHER INFORMATION: mouse TGR18 G-protein coupled receptor (GPCR)
US-09-891-138A-1
```

```
Query Match          100.0%; Score 1543; DB11; Length 1543;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 GCTCCTGGCAGAGTTTTCTGTCTGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60
        |||
Db      1 GCTCCTGGCAGAGTTTTCTGTCTGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60

Qy     61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120
        |||
Db     61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120

Qy    121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACGTGGTGTTCGGCTACCTCTT 180
        |||
Db    121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACGTGGTGTTCGGCTACCTCTT 180

Qy    181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTAACCTTTCCATCTCTGACTT 240
        |||
Db    181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTAACCTTTCCATCTCTGACTT 240

Qy    241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300
        |||
Db    241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300

Qy    301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360
        |||
Db    301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360

Qy    361 CCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA 420
        |||
Db    361 CCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA 420

Qy    421 ACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCCTTAGT 480
        |||
Db    421 ACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCCTTAGT 480
```

Qy	481	GACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAG	540
Db	481	GACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAG	540
Qy	541	TAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTG	600
Db	541	TAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTG	600
Qy	601	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Db	601	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Qy	661	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Db	661	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Qy	721	CCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Db	721	CCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Qy	781	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Db	781	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Qy	841	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGGCCATCAA	900
Db	841	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGGCCATCAA	900
Qy	901	TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAG	960
Db	901	TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAG	960
Qy	961	ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT	1020
Db	961	ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT	1020
Qy	1021	CAGCCAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTAACTAAGTAAACC	1080
Db	1021	CAGCCAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTAACTAAGTAAACC	1080
Qy	1081	ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCAGGGCTGGAGTACAAGCTG	1140
Db	1081	ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCAGGGCTGGAGTACAAGCTG	1140
Qy	1141	GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG	1200
Db	1141	GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG	1200
Qy	1201	AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA	1260
Db	1201	AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA	1260
Qy	1261	AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAAGACCTCTTTT	1320
Db	1261	AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAAGACCTCTTTT	1320

Qy	1321	TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT	1380
Db	1321	TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT	1380
Qy	1381	TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTATTCTTGTAATATTAAAA	1440
Db	1381	TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTATTCTTGTAATATTAAAA	1440
Qy	1441	TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTAT	1500
Db	1441	TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTAT	1500
Qy	1501	ATTTCTTGAAAAATAACTGCTGTGCCTAAATAAATCAATATA	1543
Db	1501	ATTTCTTGAAAAATAACTGCTGTGCCTAAATAAATCAATATA	1543

RESULT 2

US-10-272-983-35

```

; Sequence 35, Application US/10272983
; Publication No. US20030148450A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Ruoping
; APPLICANT: Dang, Huong T.
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lin, I-Lin
; TITLE OF INVENTION: Human Orphan G Protein Coupled Receptors
; FILE REFERENCE: AREN0050
; CURRENT APPLICATION NUMBER: US/10/272,983
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US/09/417,044
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 60/109,213
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: 60/120,416
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/121,851
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/123,946
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: 60/123,949
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: 60/136,436
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/136,437
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/136,439
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/136,567
; PRIOR FILING DATE: 1999-05-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 1005
; TYPE: DNA
; ORGANISM: Homo sapiens

```



US-10-272-983-35

Query Match 38.4%; Score 592.4; DB 13; Length 1005;  
Best Local Similarity 75.5%; Pred. No. 2.1e-138;  
Matches 750; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      8 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 67

Qy     99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     68 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 127

Qy    159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    128 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 187

Qy    219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    188 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 247

Qy    279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    248 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 307

Qy    339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    308 ATGCCAACCTCTATACCAGCATTCTCTTCTCACTTTTATCAGCATAGATCGATACTTGA 367

Qy    399 TCATGAAGTACCCTTTCCGAGAACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    368 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 427

Qy    459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    428 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 487

Qy    519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAACCCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    488 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT 547

Qy    579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    548 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 607

Qy    639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    608 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 667

Qy    699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCCTGGCGGTTGTGATCTTCTCTA 758
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    668 CTGCTCTGCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 727

Qy    759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    728 TGCTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 787
```



; ORGANISM: Homo sapiens  
US-10-393-807-35

Query Match 38.4%; Score 592.4; DB 13; Length 1005;  
Best Local Similarity 75.5%; Pred. No. 2.1e-138;  
Matches 750; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      8 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 67

Qy     99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     68 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 127

Qy    159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    128 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 187

Qy    219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    188 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 247

Qy    279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    248 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 307

Qy    339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    308 ATGCCAACCTCTATACCAGCATTTCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 367

Qy    399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    368 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 427

Qy    459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    428 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 487

Qy    519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    488 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAGTTCTGGAGACCCCAACT 547

Qy    579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    548 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 607

Qy    639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    608 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 667

Qy    699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    668 CTGCTCTGCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 727

Qy    759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Qy	39	GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA	98
Db	50	GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA	109
Qy	99	AGTACTACCTCTCTGCATTTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA	158
Db	110	AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA	169
Qy	159	CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT	218

Db	170	TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT	229
Qy	219	TTAACCTTTTCCATCTCTGACTTTTGCTTTCCTGTGCACCCCTTCCCATCCTGATAAAGAGTT	278
Db	230	TTTTTTTT	
Db	230	TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCCTCCCCATGCTGATAAGGAGTT	289
Qy	279	ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC	338
Db	290	TTTTTTTT	
Db	290	ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC	349
Qy	339	ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC	398
Db	350		
Db	350	ATGCCAACCTCTATACCAGCATTCTCTTCTCACTTTTATCAGCATAGATCGATACTTGA	409
Qy	399	TCATGAAGTACCCTTTCCGAGAACACTTCTACAAAAGAAGGAATTTGCCATTTTAATCT	458
Db	410		
Db	410	TAATTAAGTATCCTTTCCGAGAACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT	469
Qy	459	CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT	518
Db	470		
Db	470	CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC	529
Qy	519	CTGTCCCAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCTGGAAACCTGAAC	578
Db	530		
Db	530	CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT	589
Qy	579	ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCTCTCTCTGTGA	638
Db	590		
Db	590	ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTTCCTTATTCCTCTTTTTGTGA	649
Qy	639	TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA	698
Db	650		
Db	650	TGTGTTTCTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGTTGCTA	709
Qy	699	CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTA	758
Db	710		
Db	710	CTGCTCTGCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG	769
Qy	759	TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT	818
Db	770		
Db	770	TGCTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT	829
Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	830		
Db	830	GGAAGCAGTATCAGTGCACCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTT	889
Qy	876	TGGCCTTTCTGAACAGTGCCATCAATCCCCTCTTCTACTTTCCTCATGGGAGACCATTACA	935
Db	890		
Db	890	TGGCCTTTCTGAACAGTGTCAATCAACCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA	949
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	950		
Db	950	GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT	1009
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAA-TGAGACACTTGATAAACAG	1048
Db	1010		
Db	1010	GGGCTCATGAACCTCTACTTTCATTCAGAGAAAAGTGAGGGGCTTGTGAAACAG	1063

RESULT 5

US-09-764-886-36

; Sequence 36, Application US/09764886

; Publication No. US20030139327A9

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PTZ02

; CURRENT APPLICATION NUMBER: US/09/764,886

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 88

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 36

; LENGTH: 1436

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-886-36

Query Match 38.4%; Score 592.4; DB 13; Length 1436;

Best Local Similarity 75.3%; Pred. No. 2.6e-138;

Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

```

Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      100 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 159

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      160 AGTACTACCTTTCCATTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGAAATACCA 219

Qy      159 CTGTGGTGTTTCGGCTACCTCTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      220 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTATCTCT 279

Qy      219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      280 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 339

Qy      279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      340 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 399

Qy      339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      400 ATGCCAACCTCTATACCAGCATTCCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 459

Qy      399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      460 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 519

Qy      459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      520 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 579

```



Best Local Similarity 75.3%; Pred. No. 5.1e-138;  
Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     110 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 169

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     170 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGAAATACCA 229

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     230 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 289

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     290 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 349

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     350 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 409

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     410 ATGCCAACCTCTATACCAGCATTCTCTTCTCACTTTTATCAGCATAGATCGATACTTGA 469

Qy     399 TCATGAAGTACCCTTTCCGAGAACACTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     470 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAGAAAGAGTTTGCTATTTTAATCT 529

Qy     459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     530 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 589

Qy     519 CTGTCCCAAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCTGGAAACCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     590 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT 649

Qy     579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     650 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 709

Qy     639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     710 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCCTAAAGCAGAGGAATAGGCAGGTTGCTA 769

Qy     699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     770 CTGCTCTGCCCTTGAAAAGCCTCTCAACTGGTCATCATGGCAGTGGAATCTTCTCTG 829

Qy     759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     830 TGCTTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 889

Qy     819 G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC 875
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```







; Sequence 3, Application US/09943798  
; Patent No. US20020065215A1  
; GENERAL INFORMATION:  
; APPLICANT: Glaxo Group Limited  
; TITLE OF INVENTION: Polypeptide  
; FILE REFERENCE: QG1021  
; CURRENT APPLICATION NUMBER: US/09/943,798  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-943-798-3

Query Match 8.2%; Score 126.6; DB 9; Length 1014;  
Best Local Similarity 49.9%; Pred. No. 3e-21;  
Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

Qy	60	CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT	119
Db	59	CTTTTGGAAATTGCACTGATGAAAACATCCCCTCAAGATGCACTACCTCCCTGTTATTT	118
Qy	120	ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCT	179
Db	119	ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT	178
Qy	180	TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACT	239
Db	179	TCAAAATGAGACCTTGAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATC	238
Qy	240	TTGCTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA	296
Db	239	TGCTGTATCTGACCAGCCTCCCCTTCTGATTCACTACTATGCCAGTGGCGAAACTGGA	298
Qy	297	CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA	356
Db	299	TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCATTTCAACCTGTATAGCA	358
Qy	357	GCATCCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC	416
Db	359	GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTACCCAATGA	418
Qy	417	GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAACTCTCGCTGGCTGTCTGGGCCT	476
Db	419	GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA	478
Qy	477	TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGG	536
Db	479	TTTCACTGGTAGCTGTCACTCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA	538
Qy	537	GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC	596
Db	539	GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACTCAATACTATTAAGTGGT	592
Qy	597	TCTGCCTGACTTTGTTGGGCTTCCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA	656

```

      | | | | | | | | | | | | | | | | | | | | | |
Db      593 ACAACCTGATTTTGGACTGCAACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652
Qy      657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      | | | | | | | | | | | | | | | | | | | | | |
Db      653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAAGTACAGCTGCCTTAAGCAGA 712
Qy      717 AACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776
      | | | | | | | | | | | | | | | | | | | | | |
Db      713 AAGCACGAAGGCTAACCAATTCTGCTACTCCTTGCAATTTACGTATGTTTTTACCCTTCC 772
Qy      777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811
      | | | | | | | | | | | | | | | | | | | | | |
Db      773 ATATCTTGAGGGTCATTTCGGATCGAATCTCGCCTG 807

```

RESULT 9

US-09-885-453-2

```

; Sequence 2, Application US/09885453
; Publication No. US20030088080A1
; GENERAL INFORMATION:
; APPLICANT: Communi, Didier
; TITLE OF INVENTION: RECEPTOR GPCRx10
; FILE REFERENCE: 9409/2082
; CURRENT APPLICATION NUMBER: US/09/885,453
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/885,453
; PRIOR FILING DATE: 2001-06-21
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DNA nucleotide sequence
; LOCATION: (1)..(1014)
; OTHER INFORMATION: GPCRx10 DNA sequence
US-09-885-453-2

```

```

Query Match          8.2%; Score 126.6; DB 11; Length 1014;
Best Local Similarity 49.9%; Pred. No. 3e-21;
Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

```

```

Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      | | | | | | | | | | | | | | | | | | | | | |
Db      59 CTTTTGGAAATTGCACTGATGAAAACATCCCCTCAAGATGCACTACCTCCCTGTTATTT 118
Qy      120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCT 179
      | | | | | | | | | | | | | | | | | | | | | |
Db      119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178
Qy      180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTTTCCATCTCTGACT 239
      | | | | | | | | | | | | | | | | | | | | | |
Db      179 TCAAAATGAGACCTTGGAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATC 238
Qy      240 TTGCTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296

```

```

      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      239 TGCTGTATCTGACCAGCCTCCCCTTCCTGATTCACTACTATGCCAGTGGCGAAAACTGGA 298
Qy      297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      || ||||| || | |||| | ||| | | ||||| || | ||
Db      299 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCATTTCAACCTGTATAGCA 358
Qy      357 GCATCCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      ||||| ||||| || | |||| | || | | || | |
Db      359 GCATCCTCTTCCTCACCTGTTTCAGCATCTCCGCTACTGTGTGATCATTCACCCAATGA 418
Qy      417 GAGAACAACCTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476
      |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      419 GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478
Qy      477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCAAAAGAAGAGG 536
      |   ||| || | || || | | |||| | | | | |
Db      479 TTTCACTGGTAGCTGTCACTCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538
Qy      537 GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596
      |   ||| |||| | |||| | | |||| |||| || | |
Db      539 GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACTCAATACTATTAAGTGGT 592
Qy      597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656
      |   |||| |||| |   |   |   |   |   |   |   |   |
Db      593 ACAACCTGATTTTGAAGTGAACACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652
Qy      657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      || |   | || |   || |   | |||| | | |   |||
Db      653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACCTGACAGCTGCCTTAAGCAGA 712
Qy      717 AACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776
      || ||| || || | ||| | | || | || || |||| |
Db      713 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTTACGTATGTTTTTTACCCCTCC 772
Qy      777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811
      ||||| || | | ||||| || |||||
Db      773 ATATCTTGAGGGTCATTCGGATCGAATCTCGCCTG 807

```

# RESULT 10

US-10-321-807-27

; Sequence 27, Application US/10321807

; Publication No. US20030166148A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Rupong

; APPLICANT: Dang, Huong T.

; APPLICANT: Lowitz, Kevin P.

; TITLE OF INVENTION: No. US20030166148A1-Endogenous, Constitutively Activated Human G Protein-Coupled

; TITLE OF INVENTION: Receptors

; FILE REFERENCE: AREN0086

; CURRENT APPLICATION NUMBER: US/10/321,807

; CURRENT FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: US/09/714,008

; PRIOR FILING DATE: 2000-11-16

; PRIOR APPLICATION NUMBER: 09/170,496

```

; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US99/23938
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/166,088
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 60/166,099
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 60/166,369
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 60/171,902
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/171,901
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/171,900
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/181,749
; PRIOR FILING DATE: 2000-02-11
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-321-807-27

```

```

Query Match          8.2%; Score 126.6; DB 13; Length 1014;
Best Local Similarity 49.9%; Pred. No. 3e-21;
Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

```

```

Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      ||| || ||||| | | | | | || | | ||||| ||| |||
Db      59 CTTTTGGAAATTGCACTGATGAAAACATCCCACCTCAAGATGCACTACCTCCCTGTTATTT 118

Qy     120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCT 179
      ||| || ||| | | ||| | | || |||| | || | | | ||| | |
Db     119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178

Qy     180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTTTCCATCTCTGACT 239
      || |||| | |||| ||||| | | | | | |||| | | | | |
Db     179 TCAAAATGAGACCTTGAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATC 238

Qy     240 TTGCTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296
      | | | ||| || ||| || ||||| | | ||||| | || || |||
Db     239 TGCTGTATCTGACCAGCCTCCCCTTCCTGATTCACTACTATGCCAGTGGCGAAACTGGA 298

Qy     297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      || ||||| | | |||| | ||| | | || | ||||| || | ||
Db     299 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCACCTGTATAGCA 358

Qy     357 GCATCCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      ||||| ||||| || | |||| | || | | | || | |||| |
Db     359 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTCACCCAATGA 418

Qy     417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476
      | | | ||| | | ||| || || | | | | | || ||| |
Db     419 GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478

```

```

Qy      477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAGAAGAGG 536
      |      ||| || | || ||| | | || ||| || | | |
Db      479 TTTCACTGGTAGCTGTCAATTCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538

Qy      537 GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596
      |      ||| |||| | |||| | | |||| |||| ||| | |
Db      539 GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACTCAATACTATTAAGTGGT 592

Qy      597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656
      | ||||| |||| | | | ||| | || | | | | | |
Db      593 ACAACCTGATTTTGACTGCAACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652

Qy      657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      || | | || | | | | | |||| | | | | | | |
Db      653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAAGTACAGCTGCCTTAAGCAGA 712

Qy      717 AACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776
      ||| ||| ||| | ||| | | ||| || || |||| |
Db      713 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTTACGTATGTTTTTTACCCCTCC 772

Qy      777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811
      ||||| || | | ||||| || |||||
Db      773 ATATCTTGAGGGTCATTTCGGATCGAATCTCGCCTG 807

```

# RESULT 11

US-10-270-144-1

; Sequence 1, Application US/10270144

; Publication No. US20030049790A1

; GENERAL INFORMATION:

; APPLICANT: WEI, Ming-Hui et al

; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED

; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR

; TITLE OF INVENTION: PROTEINS, AND USES THEREOF

; FILE REFERENCE: CL000750CON

; CURRENT APPLICATION NUMBER: US/10/270,144

; CURRENT FILING DATE: 2002-10-15

; PRIOR APPLICATION NUMBER: 60/205,196

; PRIOR FILING DATE: 2000-05-18

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 1014

; TYPE: DNA

; ORGANISM: Human

US-10-270-144-1

Query Match 8.2%; Score 126.6; DB 15; Length 1014;

Best Local Similarity 49.9%; Pred. No. 3e-21;

Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

```

Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      ||| || |||| | | | ||| ||||| ||| |||
Db      59 CTTTTGGAAATTGCACTGATGAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTT 118

Qy      120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCT 179

```





```
; APPLICANT: Cutler, Gene
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: No. US20030082585A1el Receptors
; FILE REFERENCE: 018781-008410US
; CURRENT APPLICATION NUMBER: US/10/188,405
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US 60/302,800
; PRIOR FILING DATE: 2001-07-03
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human TGR164
US-10-188-405-7
```

```
Query Match          8.2%; Score 126.6; DB 15; Length 1014;
Best Local Similarity 49.9%; Pred. No. 3e-21;
Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;
```

```
Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      ||| || ||||| | | | | | ||| ||| |||
Db      59 CTTTGGAAATTGCACTGATGAAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTT 118

Qy     120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCT 179
      ||| || ||| | | ||| | | || ||| ||| | | ||| | |
Db     119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178

Qy     180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACT 239
      || |||| |||| ||||| || | | | |||| || | | ||
Db     179 TCAAAATGAGACCTTGAAGAGCAGCACCATCATATGCTGAACCTGGCCTGCACAGATC 238

Qy     240 TTGCTTTCTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296
      | | | || || || || ||||| | . ||||| | || || |||
Db     239 TGCTGTATCTGACCAGCCTCCCCTTCTGATTCACTACTATGCCAGTGGCGAAAACCTGGA 298

Qy     297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      || ||||| | | |||| | ||| | || || ||||| || | ||
Db     299 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCATTTCAACCTGTATAGCA 358

Qy     357 GCATCCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      ||||| ||||| || | |||| ||| || | || | |||| |
Db     359 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTACCCAATGA 418

Qy     417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476
      | | | || || | || || || || || | || || || |
Db     419 GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478

Qy     477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATTCTGTCCCAAAAGAAGAGG 536
      | ||| || | || || | || |||| || | | | |
Db     479 TTTCACTGGTAGCTGTCAATCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538

Qy     537 GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596
      | ||| |||| | |||| | |||| |||| ||| | |
Db     539 GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACCTCAATACTATTAAGTGGT 592
```

Qy 597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 593 ACAACCTGATTTTGACTGCAACTACTTTCTGCCCTCCCCTTGGTGATAGTGACACTTTGCT 652  
 Qy 657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACGACAGCTGCCCTAAGCAGA 712  
 Qy 717 AACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 713 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTACGTATGTTTTTTACCCTTCC 772  
 Qy 777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 773 ATATCTTGAGGGTCATTCGGATCGAATCTCGCCTG 807

# RESULT 13

US-10-079-384-13

; Sequence 13, Application US/10079384

; Publication No. US20030108986A1

; GENERAL INFORMATION:

; APPLICANT: Communi, Didier

; TITLE OF INVENTION: COMPOSITIONS AND METHODS COMPRISING G-PROTEIN COUPLED RECEPTORS

; FILE REFERENCE: 9409/2132

; CURRENT APPLICATION NUMBER: US/10/079,384

; CURRENT FILING DATE: 2002-02-20

; PRIOR APPLICATION NUMBER: US 09/885,453

; PRIOR FILING DATE: 2001-06-20

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 13

; LENGTH: 1014

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1014)

; OTHER INFORMATION:

US-10-079-384-13

Query Match 8.2%; Score 126.6; DB 15; Length 1014;

Best Local Similarity 49.9%; Pred. No. 3e-21;

Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

Qy 60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 59 CTTTGGAAATTGCACTGATGAAACATCCCCTCAAGATGCACTACCTCCCTGTTATTT 118  
 Qy 120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCT 179  
 | | | | | | | | | | | | | | | | | | | | | |  
 Db 119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178  
 Qy 180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTAACCTTTCCATCTCTGACT 239  
 | | | | | | | | | | | | | | | | | | | | | |



; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 646  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-225-567A-646

Query Match 8.2%; Score 126.6; DB 15; Length 1014;  
Best Local Similarity 49.9%; Pred. No. 3e-21;  
Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

```
Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      ||| || ||||| | | | | | | | | | | | | | |
Db      59 CTTTTGGAAATTGCACTGATGAAACATCCCCTCAAGATGCACTACCTCCCTGTTATTT 118

Qy     120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCT 179
      ||| || ||| | | ||| | | || ||| ||| | | | | |
Db     119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178

Qy     180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACT 239
      || |||| |||| |||| || || || || || || || ||
Db     179 TCAAAATGAGACCTTGGAAGAGCAGCACCATCATATGCTGAACCTGGCCTGCACAGATC 238

Qy     240 TTGCTTTCCTGTGCACCCTTCCCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296
      | | | ||| || ||| ||| ||||| | | ||||| | | || |||
Db     239 TGCTGTATCTGACCAGCCTCCCCTTCTGATTCACTACTATGCCAGTGGCGAAAACCTGGA 298

Qy     297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      || ||||| || | |||| | ||| | | || ||||| || | ||
Db     299 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCA 358

Qy     357 GCATCCTCTTCCTCACTTTTATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      ||||| ||||| || | |||| ||| || | || | ||| |
Db     359 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCGGCTACTGTGTGATCATTCACCCAATGA 418

Qy     417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCCT 476
      | | | ||| || | || ||| || || || | || ||| |
Db     419 GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478

Qy     477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCAAAAGAAGAGG 536
      | ||| || | || || | || ||| || | | | |
Db     479 TTTCACTGGTAGCTGTCAATCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538

Qy     537 GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596
      | ||| |||| | |||| | |||| |||| ||| | |
Db     539 GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACCTCAATACTATTAAGTGGT 592

Qy     597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656
      | |||| |||| | | ||| ||| ||| | | | | |
Db     593 ACAACCTGATTTTGAAGTGAACACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652

Qy     657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      || | || | || || || |||| | | | | |
Db     653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACCTGACAGCTGCCTTAAGCAGA 712
```



Qy 297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356  
 || ||||| | | ||| | || | ||||| || | ||  
 Db 352 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCA 411

Qy 357 GCATCCTCTTCCTCACCTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416  
 ||||| ||||| | | |||| | || | |||| |  
 Db 412 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTCACCCAATGA 471

Qy 417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476  
 | | | | | | | | | | | | | | | | | |  
 Db 472 GCTGCTTTTCCATTACAAAACTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 531

Qy 477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATTCTGTCCCAAAAGAAGAGG 536  
 | ||| || | | ||| | || |||| | | | |  
 Db 532 TTTCACTGGTAGCTGTCAATTCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 591

Qy 537 GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596  
 | ||| |||| | |||| | |||| |||| ||| | |  
 Db 592 GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACTCAATACTATTAAGTGGT 645

Qy 597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656  
 | |||| |||| | | ||| | || | | | |  
 Db 646 ACAACCTGATTTTGACTGCAACTACTTTCGCCTCCCCTTGGTGATAGTGACACTTTGCT 705

Qy 657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716  
 | | | | | | | | | | | | | | | | | |  
 Db 706 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAAGTACAGCTGCCTTAAGCAGA 765

Qy 717 AACCCCAACGCCTGGTGGTCTGGCGGTGTGTATCTTCTCTATACTCTTCACACCCTATC 776  
 || | | | | | | | | | | | | | | | |  
 Db 766 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTTACGTATGTTTTTACCCTTCC 825

Qy 777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811  
 ||||| || | | ||||| || |||||  
 Db 826 ATATCTTGAGGGTCATTTCGGATCGAATCTCGCCTG 860

Search completed: December 14, 2003, 17:43:01  
 Job time : 519 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2003, 11:13:44 ; Search time 5878 Seconds  
(without alignments)  
10738.951 Million cell updates/sec

Title: US-09-891-138A-1  
Perfect score: 1543  
Sequence: 1 gctcctggcagagttttctg.....tgcctaaataaatcaatata 1543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*

28: em\_un:\*  
 29: em\_vi:\*  
 30: em\_htg\_hum:\*  
 31: em\_htg\_inv:\*  
 32: em\_htg\_other:\*  
 33: em\_htg\_mus:\*  
 34: em\_htg\_pln:\*  
 35: em\_htg\_rod:\*  
 36: em\_htg\_mam:\*  
 37: em\_htg\_vrt:\*  
 38: em\_sy:\*  
 39: em\_htgo\_hum:\*  
 40: em\_htgo\_mus:\*  
 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query						Description
	No.	Score	Match	Length	DB	ID	
	1	1543	100.0	1543	6	AX376573	AX376573 Sequence
	2	1533.4	99.4	1598	10	AF295367	AF295367 Mus muscu
c	3	1494.8	96.9	184306	2	AC138318	AC138318 Mus muscu
c	4	1041.6	67.5	239576	2	AC111231	AC111231 Rattus no
	5	645.8	41.9	60298	2	AC116149	AC116149 Mus muscu
c	6	633.6	41.1	60298	2	AC116149	AC116149 Mus muscu
c	7	615.8	39.9	326606	2	AC110839	AC110839 Rattus no
	8	592.4	38.4	1325	9	AF247785	AF247785 Homo sapi
	9	592.4	38.4	1380	6	AX549281	AX549281 Sequence
	10	592.4	38.4	1380	9	AF348078	AF348078 Homo sapi
	11	592.4	38.4	1449	9	BC030948	BC030948 Homo sapi
	12	592.4	38.4	1542	6	AX342665	AX342665 Sequence
	13	590.2	38.3	90343	9	AC116026	AC116026 Homo sapi
	14	590.2	38.3	132745	9	AC068647	AC068647 Homo sapi
	15	589.2	38.2	1996	6	AR035943	AR035943 Sequence
c	16	588.6	38.1	158144	2	AC016455	AC016455 Homo sapi
	17	572.8	37.1	1005	6	AX657428	AX657428 Sequence
	18	147.6	9.6	1977	5	AF031897	AF031897 Meleagris
	19	145.2	9.4	69462	2	AC101335	AC101335 Mus muscu
	20	126.6	8.2	1011	6	BD171271	BD171271 Novel G p
	21	126.6	8.2	1014	6	AX148186	AX148186 Sequence
	22	126.6	8.2	1014	6	AX379468	AX379468 Sequence
	23	126.6	8.2	1014	6	AX384211	AX384211 Sequence
	24	126.6	8.2	1014	6	AX549361	AX549361 Sequence
	25	126.6	8.2	1014	6	AX593341	AX593341 Sequence
	26	126.6	8.2	1014	6	BD144282	BD144282 Novel G-p
	27	126.6	8.2	1014	9	AB083598	AB083598 Homo sapi
	28	126.6	8.2	1014	9	AF411109	AF411109 Homo sapi
	29	126.6	8.2	1029	6	AX593346	AX593346 Sequence
	30	126.6	8.2	1081	6	AX458238	AX458238 Sequence
	31	126.6	8.2	1092	9	AF370886	AF370886 Homo sapi
	32	126.6	8.2	1179	6	AX593351	AX593351 Sequence
	33	126.6	8.2	1414	6	AX646271	AX646271 Sequence



	34	126.6	8.2	1414	9	AB065877	AB065877 Homo sapi
	35	126.6	8.2	1729	6	AX191332	AX191332 Sequence
	36	126.6	8.2	1797	6	AX593340	AX593340 Sequence
	37	126.6	8.2	9905	6	AX379470	AX379470 Sequence
c	38	126.6	8.2	67645	9	AL356486	AL356486 Human DNA
	39	126.6	8.2	156555	9	AC026756	AC026756 Homo sapi
	40	125.6	8.1	990	6	AX657422	AX657422 Sequence
	41	125	8.1	1014	6	AX305130	AX305130 Sequence
	42	125	8.1	1014	6	AX464561	AX464561 Sequence
	43	125	8.1	1560	6	AX675828	AX675828 Sequence
	44	125	8.1	1851	6	AX661905	AX661905 Sequence
	45	123.4	8.0	1265	9	HSA305372	AJ305372 Homo sapi

# ALIGNMENTS

## RESULT 1

AX376573

LOCUS AX376573 1543 bp DNA linear PAT 01-MAR-2002

DEFINITION Sequence 1 from Patent WO0200719.

ACCESSION AX376573

VERSION AX376573.1 GI:19170674

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1

AUTHORS Lin,D.C., Zhao,J., Chen,J.L. and Cutler,G.

TITLE Novel receptors

JOURNAL Patent: WO 0200719-A 1 03-JAN-2002;

Tularik Inc. (US)

FEATURES Location/Qualifiers

source 1..1543  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"

CDS 44..997  
/note="mouse TGR18 G-protein coupled receptor (GPCR)"  
/codon\_start=1  
/protein\_id="CAD26816.1"  
/db\_xref="GI:19170675"  
/translation="MAQNLSCEWNLATEAILNKYYLSAFYAIEFIFGLLGNTVVFGY  
LFCMKNWNSSNVYLFNLSISDFAFLCTLPILIKSYANDKGTYGDVLCISNRYVLHTNL  
YTSILFLTFISMDRYLLMKYPFREHFLQKKEFAILISLAVWALVTLEVLPLMTFINSV  
PKEEGSNCIDYASSGNPEHNLIYSLCLTLLGFLIPLSVMCFYYKMVFLKRRSQQQA  
TALPLDKPQRLVVLAVVIFSILFTPYHIMRNLRIASRLDSWPQGCTQKAIKSIYTLTR  
PLAFLNSAINPIFYFLMGDHYREMLISKFRQYFKSLTSFRT"

BASE COUNT 438 a 352 c 293 g 460 t

ORIGIN

Query Match 100.0%; Score 1543; DB 6; Length 1543;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTCCTGGCAGAGTTTTCTGTCTGAGACAGAAGCCGACAGCAGAAATGGCACAGAAATTTATC 60

Db	1	 GCTCCTGGCAGAGTTTCTGTCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC	60
Qy	61	TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA	120
Db	61	 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA	120
Qy	121	TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTTCGGCTACCTCTT	180
Db	121	 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTTCGGCTACCTCTT	180
Qy	181	CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT	240
Db	181	 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT	240
Qy	241	TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA	300
Db	241	 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA	300
Qy	301	TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT	360
Db	301	 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT	360
Qy	361	CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA	420
Db	361	 CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA	420
Qy	421	ACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT	480
Db	421	 ACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT	480
Qy	481	GACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAG	540
Db	481	 GACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAG	540
Qy	541	TAAGTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTG	600
Db	541	 TAAGTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTG	600
Qy	601	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Db	601	 CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Qy	661	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Db	661	 GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Qy	721	CCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Db	721	 CCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Qy	781	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Db	781	 CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Qy	841	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAA	900

```

Db      841 GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAA 900
Qy      901 TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAG 960
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAG 960
Qy      961 ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT 1020
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT 1020
Qy     1021 CAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTAACTAAGTAAACC 1080
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTAACTAAGTAAACC 1080
Qy     1081 ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTG 1140
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTG 1140
Qy     1141 GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG 1200
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG 1200
Qy     1201 AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA 1260
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA 1260
Qy     1261 AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAATTTTAAGACCTCTTTT 1320
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAATTTTAAGACCTCTTTT 1320
Qy     1321 TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT 1380
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT 1380
Qy     1381 TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTATTCTTGTAATATTAAAA 1440
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTATTCTTGTAATATTAAAA 1440
Qy     1441 TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTAT 1500
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTAT 1500
Qy     1501 ATTTCTTGAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543
      ||||||||||||||||||||||||||||||||||||||||||
Db     1501 ATTTCTTGAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543

```

RESULT 2

AF295367

LOCUS AF295367 1598 bp mRNA linear ROD 06-APR-2001

DEFINITION Mus musculus G-protein coupled receptor GPR91 mRNA, complete cds.

ACCESSION AF295367

VERSION AF295367.1 GI:12711490

KEYWORDS .

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1598)

AUTHORS Wittenberger,T., Schaller,H.C. and Hellebrand,S.

TITLE An expressed sequence tag (EST) data mining strategy succeeding in the discovery of new G-protein coupled receptors

JOURNAL J. Mol. Biol. 307 (3), 799-813 (2001)

MEDLINE 21172992

PUBMED 11273702

REFERENCE 2 (bases 1 to 1598)

AUTHORS Wittenberger,T., Schaller,C.H. and Hellebrand,S.

TITLE Direct Submission

JOURNAL Submitted (14-AUG-2000) ZMNH, Institut fur Entwicklungsneurobiologie, Martinistr. 52, Hamburg 20246, Germany

FEATURES Location/Qualifiers

source 1. .1598

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL"

/db\_xref="taxon:10090"

CDS 74. .1027

/note="orphan receptor"

/codon\_start=1

/product="G-protein coupled receptor GPR91"

/protein\_id="AAK01867.1"

/db\_xref="GI:12711491"

/translation="MAQNLSCENWLATEAILNKYYLSAFYAIEFIFGLLGNVTVVFGY LFCMKNWNSSNVYLFNLSISDFAFLCTLPLIKSYANDKGTYGDVLCISNRYVLHTNL YTSMLLLTVISMTRYLLMKYPFREHFLQKKEFAILISLAVWALVTLEVLPLMTFINSV PKEEGSNCIDYASSGNPEHNLIYSLCLTLLGFLIPLSVMCFFYYKMVFLKRRSQQA TALPLDKPQRLVVLAVVIFSILFTPYHIMRNLRIASRLDSWPQGCTQKAIKSIYTLTR PLAFNLNSAINPIFYFLMGDHYREMLISKFRQYFKSLTSFRT"

BASE COUNT 465 a 358 c 303 g 472 t

ORIGIN

Query Match 99.4%; Score 1533.4; DB 10; Length 1598;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1537; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GCTCCTGGCAGAGTTTTCTGTCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60

Db 31 GCTCCTGGCAGAGTTTTCTGTCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 90

Qy 61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120

Db 91 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 150

Qy 121 TGCAATCGAGTTCATTTTTGGACTGCTTGGAATGTCACTGTGGTGTTCGGCTACCTCTT 180

Db 151 TGCAATCGAGTTCATTTTTGGACTGCTTGGAATGTCACTGTGGTGTTCGGCTACCTCTT 210

Qy 181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT 240

Db 211 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT 270

Qy 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300

Db 271 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 330

Qy	301	TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT	360
Db	331	TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT	390
Qy	361	CCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA	420
Db	391	GCTCTTGCTCACTGTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA	450
Qy	421	ACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT	480
Db	451	ACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT	510
Qy	481	GACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATTCTGTCCCAAAAGAAGAGGGCAG	540
Db	511	GACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATTCTGTCCCAAAAGAAGAGGGCAG	570
Qy	541	TAACTGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCTCATTTACAGCCTCTG	600
Db	571	TAACTGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCTCATTTACAGCCTCTG	630
Qy	601	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Db	631	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	690
Qy	661	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Db	691	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	750
Qy	721	CCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Db	751	CCAACGCCTGGTGGTCTGGCAGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	810
Qy	781	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Db	811	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	870
Qy	841	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAA	900
Db	871	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAA	930
Qy	901	TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTTCAG	960
Db	931	TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTTCAG	990
Qy	961	ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT	1020
Db	991	ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT	1050
Qy	1021	CAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTAACTAAGTAAACC	1080
Db	1051	CAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTAACTAAGTAAACC	1110
Qy	1081	ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTG	1140
Db	1111	ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTG	1170

Qy 1141 GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG 1200  
 |||  
 Db 1171 GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG 1230  
 Qy 1201 AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA 1260  
 |||  
 Db 1231 AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA 1290  
 Qy 1261 AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTT 1320  
 |||  
 Db 1291 AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTT 1350  
 Qy 1321 TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT 1380  
 |||  
 Db 1351 TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT 1410  
 Qy 1381 TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTATTCTTGTAATATTAAAA 1440  
 |||  
 Db 1411 TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTATTCTTGTAATATTAAAA 1470  
 Qy 1441 TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAAATTAT 1500  
 |||  
 Db 1471 TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAAATTAT 1530  
 Qy 1501 ATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543  
 |||  
 Db 1531 ATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1573

# RESULT 3

AC138318/c

LOCUS AC138318 184306 bp DNA linear HTG 22-MAR-2003

DEFINITION Mus musculus chromosome 3 clone RP23-358I23 map 3, WORKING DRAFT  
 SEQUENCE, 10 unordered pieces.

ACCESSION AC138318

VERSION AC138318.3 GI:29150501

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 184306)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

TITLE Mus musculus chromosome 3, clone RP23-358I23

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 184306)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,  
 Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,  
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,  
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (25-DEC-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE 3 (bases 1 to 184306)  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,  
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,  
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (22-MAR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Mar 22, 2003 this sequence version replaced gi:28191615.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L28921  
 Center clone name: 358\_I\_23  
 ----- Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 181695 bases at least Q40  
 Consensus quality: 182410 bases at least Q30  
 Consensus quality: 182638 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Insert size: 183406; sum-of-contigs  
 Quality coverage: 12.8 in Q20 bases; agarose-fp  
 Quality coverage: 11.9 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 58928: contig of 58928 bp in length  
 \* 58929 59028: gap of 100 bp  
 \* 59029 59675: contig of 647 bp in length  
 \* 59676 59775: gap of 100 bp  
 \* 59776 64745: contig of 4970 bp in length  
 \* 64746 64845: gap of 100 bp  
 \* 64846 69474: contig of 4629 bp in length  
 \* 69475 69574: gap of 100 bp  
 \* 69575 73672: contig of 4098 bp in length  
 \* 73673 73772: gap of 100 bp  
 \* 73773 81217: contig of 7445 bp in length  
 \* 81218 81317: gap of 100 bp  
 \* 81318 97238: contig of 15921 bp in length  
 \* 97239 97338: gap of 100 bp  
 \* 97339 113368: contig of 16030 bp in length  
 \* 113369 113468: gap of 100 bp  
 \* 113469 166137: contig of 52669 bp in length  
 \* 166138 166237: gap of 100 bp  
 \* 166238 184306: contig of 18069 bp in length.

FEATURES	Location/Qualifiers
source	1. .184306 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="3" /map="3" /clone="RP23-358I23" /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature	1. .58928 /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature	59029. .59675 /note="assembly_fragment"
misc_feature	59776. .64745 /note="assembly_fragment"
misc_feature	64846. .69474 /note="assembly_fragment"
misc_feature	69575. .73672 /note="assembly_fragment"
misc_feature	73773. .81217 /note="assembly_fragment"
misc_feature	81318. .97238 /note="assembly_fragment"
misc_feature	97339. .113368 /note="assembly_fragment"
misc_feature	113469. .166137 /note="assembly_fragment"



```

misc_feature      166238. .184306
                  /note="assembly_fragment
                  clone_end:T7
                  vector_side:right"
BASE COUNT      58167 a  34380 c  35771 g  55088 t    900 others
ORIGIN

Query Match      96.9%; Score 1494.8; DB 2; Length 184306;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1496; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      46 GGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTA 105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145736 GGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTA
145677

Qy      106 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCACTGTGGT 165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145676 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCACTGTGGT
145617

Qy      166 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACT 225
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145616 GTTTGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACT
145557

Qy      226 TTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145556 TTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA
145497

Qy      286 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA 345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145496 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA
145437

Qy      346 CCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAA 405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145436 CCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAA
145377

Qy      406 GTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGC 465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145376 GTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGC
145317

Qy      466 TGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCC 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145316 TGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCC
145257

Qy      526 AAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCT 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      145256 AAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCT
145197

Qy      586 CATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTT 645

```

Db	145196		CATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTT	
	145137			
Qy	646		CTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCT	705
Db	145136		CTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCT	
	145077			
Qy	706		GCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTT	765
Db	145076		GCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCAGTTGTGATCTTCTCTATACTCTT	
	145017			
Qy	766		CACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACA	825
Db	145016		CACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACA	
	144957			
Qy	826		AGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCT	885
Db	144956		AGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCT	
	144897			
Qy	886		GAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCT	945
Db	144896		GAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCT	
	144837			
Qy	946		GATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGG	1005
Db	144836		GATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGG	
	144777			
Qy	1006		ATGCAGGTCTTCACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTT	1065
Db	144776		ATGCAGGTCTTCACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTT	
	144717			
Qy	1066		TTAACTAAGTAAACCACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGG	1125
Db	144716		TTAACTAAGTAAACCACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGG	
	144657			
Qy	1126		CTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTT	1185
Db	144656		CTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTT	
	144597			
Qy	1186		ATACCCAGAGTATGGAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACT	1245
Db	144596		ATACCCAGAGTATGGAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACT	
	144537			
Qy	1246		GAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAAT	1305

Db 144536 GAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAAT  
144477

Qy 1306 TTTAAGACCTCTTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTG 1365  
|||||

Db 144476 TTTAAGACCTCTTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTG  
144417

Qy 1366 AATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATT 1425  
|||||

Db 144416 AATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATT  
144357

Qy 1426 CTTGTAATATTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTT 1485  
|||||

Db 144356 CTTGTAATATTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTT  
144297

Qy 1486 CTATTTGAAAATTATATTTCTTGAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543  
|||||

Db 144296 CTATTTGAAAATTATATTTCTTGAAAAATAACTGCTGTGCCTAAATAAATCAATATA 144239

#### RESULT 4

AC111231/c

LOCUS AC111231 239576 bp DNA linear HTG 13-MAY-2003

DEFINITION Rattus norvegicus clone CH230-96013, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 2 unordered pieces.

ACCESSION AC111231

VERSION AC111231.7 GI:30578486

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 239576)

AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidas,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,  
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 239576)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 3 (bases 1 to 239576)  
 AUTHORS Rat Genome Sequencing Consortium.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On May 13, 2003 this sequence version replaced gi:24819079.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GLVO  
Center clone name: CH230-96013

----- Summary Statistics

Assembly program: Atlas 3.0;  
Consensus quality: 213738 bases at least Q40  
Consensus quality: 217471 bases at least Q30  
Consensus quality: 220066 bases at least Q20  
Estimated insert size: 227472; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 236521: contig of 236521 bp in length  
\* 236522 236621: gap of unknown length  
\* 236622 239576: contig of 2955 bp in length.

FEATURES                      Location/Qualifiers  
    source                    1. .239576  
                              /organism="Rattus norvegicus"  
                              /mol\_type="genomic DNA"  
                              /db\_xref="taxon:10116"  
                              /clone="CH230-96013"  
    misc\_feature            157219. .158900  
                              /note="wgs\_contig"  
    misc\_feature            206334. .207349  
                              /note="wgs\_contig"  
BASE COUNT    67727 a 42048 c 43435 g 68312 t 18054 others  
ORIGIN

Query Match                      67.5%;    Score 1041.6;    DB 2;    Length 239576;  
Best Local Similarity    85.2%;    Pred. No. 1.3e-221;  
Matches 1287;    Conservative    0;    Mismatches 194;    Indels    29;    Gaps    10;

Qy            46 GGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTA 105  
              |||||  
Db            92574 GGCACAGAATTTATCTTGTGAAAATTGGCTGGCATTAGAGAATATTTTAAAAAGTACTA 92515  
  
Qy            106 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCACCTGTGGT 165  
              |||||  
Db            92514 CCTCTCTGCATTTTATGGGATCGAGTTCATTGTTGGAATGCTTGGCAATTTACCGTGGT 92455  
  
Qy            166 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCT 225  
              |||||  
Db            92454 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGTAGCAACGTCTATCTTTCAACCT 92395

Qy 226 TTCCATCTCTGACTTTGCTTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA 285  
 |||||  
 Db 92394 TTCCATCTCTGACCTTGCTTTTCTGTGCACGCTTCCCATGCTGATAAGGAGTTACGCCAC 92335  
 Qy 286 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA 345  
 || || |||||  
 Db 92334 TGGGAACCTGGACCTATGGAGATGTTCTCTGCATAAGCAACCGTTATGTGCTTCATGCCAA 92275  
 Qy 346 CCTCTACACCAGCATCCTCTTCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAA 405  
 |||||  
 Db 92274 CCTCTACACCAGCATCCTTTTCTCACTTTCATTAGCATAGACCGATATCTGCTCATGAA 92215  
 Qy 406 GTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGC 465  
 || |||||  
 Db 92214 GTTCCCTTTCCGAGAACACATTCTACAAAAGAAGGAATTTGCCATTTTAATCTCCCTGGC 92155  
 Qy 466 TGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCC 525  
 ||||| |||||  
 Db 92154 TGTCTGGGTCTTAGTGACCTTAGAAGTTCTACCTATGCTCACGTTTATCACTTCCACCCC 92095  
 Qy 526 AAAAGAAGAGGGCAGTAACCTGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCT 585  
 || |||| |||| | ||| |||||  
 Db 92094 AATAGAAAAGGGCGACAGCTGTGTGACTATGCAAGTTCTGGAAACCTAAATACAGTCT 92035  
 Qy 586 CATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTT 645  
 ||||| ||||| ||||| ||||| ||||| |||||  
 Db 92034 CATTTACAGCCTGTGCCTGACTTTGCTGGGCTTCCTCATTCCTCTGTCTGTAATGTGCTT 91975  
 Qy 646 CTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCT 705  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91974 CTTCTACTACAAAATGGTAGTCTTCCTAAAGAAGAGGAGCCAGCAGCAGGCAACTGTGCT 91915  
 Qy 706 GCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTT 765  
 | ||| ||||| | ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91914 ATCGCTGAACAAACCTCTGCGCCTGGTGGTCTGGCAGTGGTGTGATCTTCTCTGTACTCTT 91855  
 Qy 766 CACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACA 825  
 |||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91854 TACACCTTACCATATCATGCGCAATGTGAGGATTGCCTCACGCTTGGATAGCTGGCCACA 91795  
 Qy 826 AGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCT 885  
 ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91794 GGGATGTTCCCAGAAGGCCATCAAATGCTTATACATCCTGACCAGACCTCTGGCCTTTCT 91735  
 Qy 886 GAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCT 945  
 ||||| |||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91734 GAACAGTGCTGTCAACCCCATCTTCTACTTCCTTGTGGGAGACATTTTACAGAGACATGCT 91675  
 Qy 946 GATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGG 1005  
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 91674 GTTTAGTAAGTTGAGACAATACTTCAAGTCCCTTACGTCCTTCAGGCTCTGACCT----A 91619  
 Qy 1006 ATGCAGGTCTTCACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTT 1065  
 ||| ||||| |||| ||| |||| | || ||||| ||||| |||||  
 Db 91618 ATGTAGGTCTTCACTGAGCCAGAATAAGACTC-----AACTCTGCAGTTGAGTT 91570  
 Qy 1066 TTAAC TAAGTAAACCACCATTTCTAGGCTTTAGC-TTTCCACCATCCTCCAACCCCCAGG 1124

Db 91569 TTGACCAAGTAGACCACCACCTCTAGGCTTTAGCGTTCCCACCATCCTCCAACCCTGAGT 91510  
 Qy 1125 GCTGGAGTACAAGCTGGGTCCACATGAATCAGAAG-GCAGCTCTCTGTTCTGATTTTAGG 1183  
 Db 91509 GCTAGAGCACAACTGGGCACACATGAATCAGAAGAGCAACCATCTGTCCCGATTTTAGG 91450  
 Qy 1184 TTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAA 1243  
 Db 91449 CTGTACCCAGAGTATGG-AAAAATGAGGCCCCAGAAAGCATTGACATCTTACATAAGAA 91391  
 Qy 1244 CTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGA 1303  
 Db 91390 CTGAACAAAAGAAAACCTGATGTTGTCAATATTGGACACTTAAGATCCAAGGCGTTGGAG 91331  
 Qy 1304 ATTTTAAGACCTCTT-TTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAATG 1362  
 Db 91330 ATTTTAAGACATCTTCTTTCTATCAGTGTAAGGAATACGAGACAGCTAGTT-CTGACA 91272  
 Qy 1363 CTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTC---TTAT 1418  
 Db 91271 CTGAATGCATTTTGTCTATTGGTCAGCTTGATAAGAATGTTTCTGAAATAGTCTCTATTAT 91212  
 Qy 1419 TTTTATTCTTGTAATATTAA-AATTTATGTGAAAAATGAATATAATTCAATGTACAACAT 1477  
 Db 91211 TTTTATTCTTGCAATATTAACTTTTATATGAATGGTGAGTAGAACTCAATGTACAACAT 91152  
 Qy 1478 TAGATTTTCTATTTGAAAATTATATTTCTTGAAAA----AATAACTGCTGTGCCTAAATA 1533  
 Db 91151 TAGCAATTATATTCAGAAAGTACATTCTTGAAAAAATGAATAACTGCAATGCCTAAATA 91092  
 Qy 1534 AATCAATATA 1543  
 Db 91091 AATCAACACA 91082

# RESULT 5

AC116149

LOCUS AC116149 60298 bp DNA linear HTG 25-MAR-2002

DEFINITION Mus musculus clone RP24-540E9, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC116149

VERSION AC116149.1 GI:19703273

KEYWORDS HTG; HTGS PHASE0.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 60298)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Mus musculus, clone RP24-540E9

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 60298)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
 Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
 Center project name: L24912  
 Center clone name: 540\_E\_9  
 -----

\* NOTE: This record contains 77 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 656: contig of 656 bp in length  
 \* 657 756: gap of 100 bp  
 \* 757 1426: contig of 670 bp in length  
 \* 1427 1526: gap of 100 bp  
 \* 1527 2210: contig of 684 bp in length  
 \* 2211 2310: gap of 100 bp  
 \* 2311 2997: contig of 687 bp in length  
 \* 2998 3097: gap of 100 bp  
 \* 3098 3786: contig of 689 bp in length  
 \* 3787 3886: gap of 100 bp  
 \* 3887 4577: contig of 691 bp in length  
 \* 4578 4677: gap of 100 bp  
 \* 4678 5357: contig of 680 bp in length  
 \* 5358 5457: gap of 100 bp



*	5458	6150: contig of 693 bp in length
*	6151	6250: gap of 100 bp
*	6251	6817: contig of 567 bp in length
*	6818	6917: gap of 100 bp
*	6918	7615: contig of 698 bp in length
*	7616	7715: gap of 100 bp
*	7716	8412: contig of 697 bp in length
*	8413	8512: gap of 100 bp
*	8513	9198: contig of 686 bp in length
*	9199	9298: gap of 100 bp
*	9299	9988: contig of 690 bp in length
*	9989	10088: gap of 100 bp
*	10089	10768: contig of 680 bp in length
*	10769	10868: gap of 100 bp
*	10869	11524: contig of 656 bp in length
*	11525	11624: gap of 100 bp
*	11625	12242: contig of 618 bp in length
*	12243	12342: gap of 100 bp
*	12343	13040: contig of 698 bp in length
*	13041	13140: gap of 100 bp
*	13141	13829: contig of 689 bp in length
*	13830	13929: gap of 100 bp
*	13930	14647: contig of 718 bp in length
*	14648	14747: gap of 100 bp
*	14748	15451: contig of 704 bp in length
*	15452	15551: gap of 100 bp
*	15552	16247: contig of 696 bp in length
*	16248	16347: gap of 100 bp
*	16348	17028: contig of 681 bp in length
*	17029	17128: gap of 100 bp
*	17129	17802: contig of 674 bp in length
*	17803	17902: gap of 100 bp
*	17903	18593: contig of 691 bp in length
*	18594	18693: gap of 100 bp
*	18694	19375: contig of 682 bp in length
*	19376	19475: gap of 100 bp
*	19476	20082: contig of 607 bp in length
*	20083	20182: gap of 100 bp
*	20183	20875: contig of 693 bp in length
*	20876	20975: gap of 100 bp
*	20976	21650: contig of 675 bp in length
*	21651	21750: gap of 100 bp
*	21751	22427: contig of 677 bp in length
*	22428	22527: gap of 100 bp
*	22528	23238: contig of 711 bp in length
*	23239	23338: gap of 100 bp
*	23339	24028: contig of 690 bp in length
*	24029	24128: gap of 100 bp
*	24129	24803: contig of 675 bp in length
*	24804	24903: gap of 100 bp
*	24904	25603: contig of 700 bp in length
*	25604	25703: gap of 100 bp
*	25704	26357: contig of 654 bp in length
*	26358	26457: gap of 100 bp
*	26458	27140: contig of 683 bp in length
*	27141	27240: gap of 100 bp
*	27241	27946: contig of 706 bp in length

*	27947	28046: gap of 100 bp
*	28047	28734: contig of 688 bp in length
*	28735	28834: gap of 100 bp
*	28835	29536: contig of 702 bp in length
*	29537	29636: gap of 100 bp
*	29637	30324: contig of 688 bp in length
*	30325	30424: gap of 100 bp
*	30425	31130: contig of 706 bp in length
*	31131	31230: gap of 100 bp
*	31231	31910: contig of 680 bp in length
*	31911	32010: gap of 100 bp
*	32011	32691: contig of 681 bp in length
*	32692	32791: gap of 100 bp
*	32792	33482: contig of 691 bp in length
*	33483	33582: gap of 100 bp
*	33583	34274: contig of 692 bp in length
*	34275	34374: gap of 100 bp
*	34375	35081: contig of 707 bp in length
*	35082	35181: gap of 100 bp
*	35182	35861: contig of 680 bp in length
*	35862	35961: gap of 100 bp
*	35962	36660: contig of 699 bp in length
*	36661	36760: gap of 100 bp
*	36761	37447: contig of 687 bp in length
*	37448	37547: gap of 100 bp
*	37548	38243: contig of 696 bp in length
*	38244	38343: gap of 100 bp
*	38344	39034: contig of 691 bp in length
*	39035	39134: gap of 100 bp
*	39135	39813: contig of 679 bp in length
*	39814	39913: gap of 100 bp
*	39914	40597: contig of 684 bp in length
*	40598	40697: gap of 100 bp
*	40698	41392: contig of 695 bp in length
*	41393	41492: gap of 100 bp
*	41493	42190: contig of 698 bp in length
*	42191	42290: gap of 100 bp
*	42291	42967: contig of 677 bp in length
*	42968	43067: gap of 100 bp
*	43068	43736: contig of 669 bp in length
*	43737	43836: gap of 100 bp
*	43837	44525: contig of 689 bp in length
*	44526	44625: gap of 100 bp
*	44626	45306: contig of 681 bp in length
*	45307	45406: gap of 100 bp
*	45407	46111: contig of 705 bp in length
*	46112	46211: gap of 100 bp
*	46212	46848: contig of 637 bp in length
*	46849	46948: gap of 100 bp
*	46949	47639: contig of 691 bp in length
*	47640	47739: gap of 100 bp
*	47740	48431: contig of 692 bp in length
*	48432	48531: gap of 100 bp
*	48532	49221: contig of 690 bp in length
*	49222	49321: gap of 100 bp
*	49322	50017: contig of 696 bp in length
*	50018	50117: gap of 100 bp

Query Match 41.9%; Score 645.8; DB 2; Length 60298;  
Best Local Similarity 84.0%; Pred. No. 2.4e-133;  
Matches 673; Conservative 0; Mismatches 127; Indels 1; Gaps 1;

Qy	51	AGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCT	110
Db	3890	AGATCTGATATCTCGCCCTGTGGTGGGAATTCTCAGGCTATCTTGAATAAGTACTACCTCT	3949
Qy	111	CTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCG	170
Db	3950	CTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCG	4009
Qy	171	GCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTTTCCA	230
Db	4010	GCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTTTCCA	4069
Qy	231	TCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATA	290
Db	4070	TCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATA	4129
Qy	291	AGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCT	350
Db	4130	AGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCT	4189
Qy	351	ACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACC	410
Db	4190	ACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACC	4249
Qy	411	CTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCT	470
Db	4250	CTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCT	4309
Qy	471	GGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAAG	530
Db	4310	GGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAAG	4369
Qy	531	AAGAGGGCAGTAACATGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCTCATTT	590
Db	4370	AAGAGGGCAGTAACATGCATCGACTATGCAAGTTCTGGAAACCTGAACACAATCTCATTT	4429
Qy	591	ACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCT	650
Db	4430	ACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCT	4489
Qy	651	ACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCAC	710

Db 4490 ACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCAC 4549  
 Qy 711 TGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACAC 770  
 ||||| ||||| ||||| ||||| |||||  
 Db 4550 TGGAC-AACCCAAACGCCTGGGGGTCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4608  
 Qy 771 CCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGAT 830  
 Db 4609 NNN 4668  
 Qy 831 GTACACAGAAGGCCATCAAAT 851  
 || ||| ||  
 Db 4669 NNNNNNNNNCGGAGATCTGAT 4689

# RESULT 6

AC116149/c

LOCUS AC116149 60298 bp DNA linear HTG 25-MAR-2002

DEFINITION Mus musculus clone RP24-540E9, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC116149

VERSION AC116149.1 GI:19703273

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 60298)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Mus musculus, clone RP24-540E9

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 60298)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
 Landers,T., Lehoczký,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L24912  
Center clone name: 540\_E\_9

-----  
\* NOTE: This record contains 77 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

* 1	656: contig of 656 bp in length
* 657	756: gap of 100 bp
* 757	1426: contig of 670 bp in length
* 1427	1526: gap of 100 bp
* 1527	2210: contig of 684 bp in length
* 2211	2310: gap of 100 bp
* 2311	2997: contig of 687 bp in length
* 2998	3097: gap of 100 bp
* 3098	3786: contig of 689 bp in length
* 3787	3886: gap of 100 bp
* 3887	4577: contig of 691 bp in length
* 4578	4677: gap of 100 bp
* 4678	5357: contig of 680 bp in length
* 5358	5457: gap of 100 bp
* 5458	6150: contig of 693 bp in length
* 6151	6250: gap of 100 bp
* 6251	6817: contig of 567 bp in length
* 6818	6917: gap of 100 bp
* 6918	7615: contig of 698 bp in length
* 7616	7715: gap of 100 bp
* 7716	8412: contig of 697 bp in length
* 8413	8512: gap of 100 bp
* 8513	9198: contig of 686 bp in length
* 9199	9298: gap of 100 bp
* 9299	9988: contig of 690 bp in length
* 9989	10088: gap of 100 bp
* 10089	10768: contig of 680 bp in length
* 10769	10868: gap of 100 bp
* 10869	11524: contig of 656 bp in length
* 11525	11624: gap of 100 bp
* 11625	12242: contig of 618 bp in length
* 12243	12342: gap of 100 bp
* 12343	13040: contig of 698 bp in length
* 13041	13140: gap of 100 bp
* 13141	13829: contig of 689 bp in length

*	13830	13929: gap of 100 bp
*	13930	14647: contig of 718 bp in length
*	14648	14747: gap of 100 bp
*	14748	15451: contig of 704 bp in length
*	15452	15551: gap of 100 bp
*	15552	16247: contig of 696 bp in length
*	16248	16347: gap of 100 bp
*	16348	17028: contig of 681 bp in length
*	17029	17128: gap of 100 bp
*	17129	17802: contig of 674 bp in length
*	17803	17902: gap of 100 bp
*	17903	18593: contig of 691 bp in length
*	18594	18693: gap of 100 bp
*	18694	19375: contig of 682 bp in length
*	19376	19475: gap of 100 bp
*	19476	20082: contig of 607 bp in length
*	20083	20182: gap of 100 bp
*	20183	20875: contig of 693 bp in length
*	20876	20975: gap of 100 bp
*	20976	21650: contig of 675 bp in length
*	21651	21750: gap of 100 bp
*	21751	22427: contig of 677 bp in length
*	22428	22527: gap of 100 bp
*	22528	23238: contig of 711 bp in length
*	23239	23338: gap of 100 bp
*	23339	24028: contig of 690 bp in length
*	24029	24128: gap of 100 bp
*	24129	24803: contig of 675 bp in length
*	24804	24903: gap of 100 bp
*	24904	25603: contig of 700 bp in length
*	25604	25703: gap of 100 bp
*	25704	26357: contig of 654 bp in length
*	26358	26457: gap of 100 bp
*	26458	27140: contig of 683 bp in length
*	27141	27240: gap of 100 bp
*	27241	27946: contig of 706 bp in length
*	27947	28046: gap of 100 bp
*	28047	28734: contig of 688 bp in length
*	28735	28834: gap of 100 bp
*	28835	29536: contig of 702 bp in length
*	29537	29636: gap of 100 bp
*	29637	30324: contig of 688 bp in length
*	30325	30424: gap of 100 bp
*	30425	31130: contig of 706 bp in length
*	31131	31230: gap of 100 bp
*	31231	31910: contig of 680 bp in length
*	31911	32010: gap of 100 bp
*	32011	32691: contig of 681 bp in length
*	32692	32791: gap of 100 bp
*	32792	33482: contig of 691 bp in length
*	33483	33582: gap of 100 bp
*	33583	34274: contig of 692 bp in length
*	34275	34374: gap of 100 bp
*	34375	35081: contig of 707 bp in length
*	35082	35181: gap of 100 bp
*	35182	35861: contig of 680 bp in length
*	35862	35961: gap of 100 bp

Query Match 41.1%; Score 633.6; DB 2; Length 60298;  
Best Local Similarity 97.6%; Pred. No. 1.2e-130;  
Matches 664; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

[illegible]

Db 36599 AAGGAATTTGCCATTTTAAATCTCGCTGGCTGTCTGGGCTTTAGTGACCTTAGAAGTTCTA 36540  
 Qy 497 CCCATGCTCACTTTTCATCAATTCTGTCCCAAAGAAGAGGGCAGTAACTGCATCGACTAT 556  
 |||||  
 Db 36539 CCCATGCTCACTTTTCATCAATTCTGTCCCAAAGAAGAGGGCAGTAACTGCATCGACTAT 36480  
 Qy 557 GCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGC 616  
 |||||  
 Db 36479 GCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGC 36420  
 Qy 617 TTCCTAATTCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAG 676  
 |||||  
 Db 36419 TTCCTAATTCTCTCTCTGTGATGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAG 36360  
 Qy 677 AGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTC 736  
 |||||  
 Db 36359 AGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTC 36300  
 Qy 737 CTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGG 796  
 |||||  
 Db 36299 CTGGCAGTTGTGATCTTCTCTATACTCTTCACACCCTATCATATCATGCGCAATTTGAGG 36240  
 Qy 797 ATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATA 856  
 |||||  
 Db 36239 ATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAAGGCCATCAAATCTATA 36180  
 Qy 857 TACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTC 916  
 |||||  
 Db 36179 TACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTC 36120  
 Qy 917 CTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCC 976  
 |||||  
 Db 36119 CTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCC 36060  
 Qy 977 CTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAATGAGACA 1036  
 |||||  
 Db 36059 CTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAATGAGACA 36000  
 Qy 1037 CTTGATAAACAGTGCTGTGC 1056  
 || | || | ||  
 Db 35999 CTGAGAATCCACCACAGGGC 35980

# RESULT 7

AC110839/c

LOCUS AC110839 326606 bp DNA linear HTG 11-OCT-2002

DEFINITION Rattus norvegicus clone CH230-208A12, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 25 unordered pieces.

ACCESSION AC110839

VERSION AC110839.4 GI:23820318

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 326606)



AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 326606)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (16-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 326606)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Oct 11, 2002 this sequence version replaced gi:21739250.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GRKD

Center clone name: CH230-208A12

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 242752 bases at least Q40

Consensus quality: 250821 bases at least Q30

Consensus quality: 254983 bases at least Q20

Estimated insert size: 244968; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 25 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 10356: contig of 10356 bp in length

\* 10357 10456: gap of unknown length

\* 10457 15819: contig of 5363 bp in length

\* 15820 15919: gap of unknown length

\* 15920 245368: contig of 229449 bp in length

\* 245369 245468: gap of unknown length

\* 245469 272041: contig of 26573 bp in length

\* 272042 272141: gap of unknown length

\* 272142 276368: contig of 4227 bp in length

\* 276369 276468: gap of unknown length

\* 276469 282159: contig of 5691 bp in length

\* 282160 282259: gap of unknown length

\* 282260 283432: contig of 1173 bp in length  
 \* 283433 283532: gap of unknown length  
 \* 283533 284633: contig of 1101 bp in length  
 \* 284634 284733: gap of unknown length  
 \* 284734 285764: contig of 1031 bp in length  
 \* 285765 285864: gap of unknown length  
 \* 285865 287082: contig of 1218 bp in length  
 \* 287083 287182: gap of unknown length  
 \* 287183 288399: contig of 1217 bp in length  
 \* 288400 288499: gap of unknown length  
 \* 288500 289828: contig of 1329 bp in length  
 \* 289829 289928: gap of unknown length  
 \* 289929 291274: contig of 1346 bp in length  
 \* 291275 291374: gap of unknown length  
 \* 291375 293018: contig of 1644 bp in length  
 \* 293019 293118: gap of unknown length  
 \* 293119 294732: contig of 1614 bp in length  
 \* 294733 294832: gap of unknown length  
 \* 294833 296078: contig of 1246 bp in length  
 \* 296079 296178: gap of unknown length  
 \* 296179 297942: contig of 1764 bp in length  
 \* 297943 298042: gap of unknown length  
 \* 298043 299812: contig of 1770 bp in length  
 \* 299813 299912: gap of unknown length  
 \* 299913 301595: contig of 1683 bp in length  
 \* 301596 301695: gap of unknown length  
 \* 301696 304787: contig of 3092 bp in length  
 \* 304788 304887: gap of unknown length  
 \* 304888 306249: contig of 1362 bp in length  
 \* 306250 306349: gap of unknown length  
 \* 306350 307801: contig of 1452 bp in length  
 \* 307802 307901: gap of unknown length  
 \* 307902 309454: contig of 1553 bp in length  
 \* 309455 309554: gap of unknown length  
 \* 309555 314110: contig of 4556 bp in length  
 \* 314111 314210: gap of unknown length  
 \* 314211 326606: contig of 12396 bp in length.

FEATURES	Location/Qualifiers
source	1. .326606 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-208A12"
misc_feature	1. .1742 /note="wgs_end_extension clone_end:Sp6"
misc_feature	complement(4245. .5082) /note="clone_boundary clone_end:Sp6 site:EcoRI end sequence:RWBKN06TVB"
misc_feature	10457. .12850 /note="wgs_contig"
misc_feature	15920. .16991 /note="wgs_contig"
misc_feature	complement(220129. .221101) /note="clone_boundary"

```

clone_end:T7
site:EcoRI
end_sequence:RWBKN06TJB"
misc_feature 241580. .242749
/note="wgs_end_extension
clone_end:T7"
misc_feature 243833. .245368
/note="wgs_end_extension
clone_end:T7"

```

BASE COUNT 81699 a 50290 c 51837 g 74097 t 68683 others  
ORIGIN

Query Match 39.9%; Score 615.8; DB 2; Length 326606;  
Best Local Similarity 89.0%; Pred. No. 1.2e-126;  
Matches 665; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 46 GGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTA 105  
|||||  
Db 242326 GGCACAGAATTTATCTTGTGAAAATTGGCTGGCATTAGAGAATATTTTGAAAAGTACTA 125  
242267

Qy 106 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCACTGTGGT 165  
|||||  
Db 242266 CCTCTCTGCATTTTATGGGATCGAGTTCATTGTTGGAATGCTTGGCAATTTACCGTGGT 165  
242207

Qy 166 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACT 225  
|||||  
Db 242206 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGTAGCAACGTCTATCTCTTCACT 225  
242147

Qy 226 TTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA 285  
|||||  
Db 242146 TTCCATCTCTGACCTTGCCTTTCCTGTGCACGCTTCCCATGCTGATAAGGAGTTACGCCAC 285  
242087

Qy 286 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA 345  
|| ||  
Db 242086 TGGGAACCTGGACCTATGGAGATGTTCTCTGCATAAGCAACCGTTATGTGCTTCATGCCAA 345  
242027

Qy 346 CCTCTACACCAGCATCCTTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAA 405  
|||||  
Db 242026 CCTCTACACCAGCATCCTTTCCTCACTTTTCATTAGCATAGACCGATATCTGCTCATGAA 405  
241967

Qy 406 GTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGC 465  
||  
Db 241966 GTTCCCTTTCCGAGAACACATTCTACAAAAGAAGGAATTTGCCATTTTAATCTCCCTGGC 465  
241907

Qy 466 TGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCC 525  
|||||  
Db 241906 TGTCTGGGTCTTAGTGACCTTAGAAGTTCTACCTATGCTCACGTTTATCACTTCCACCCC 525  
241847

Qy 526 AAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCT 585

```

      || |||| |||| | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241846 AATAGAAAAGGGGCGACAGCTGTGTGCGACTATGCAAGTTCGGAAACCCTAAATACAGTCT
241787

Qy      586 CATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTT 645
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241786 CATTTACAGCCTGTGCCTGACTTTGCTGGGCTTCCTCATTCTCTGTCTGTAATGTGCTT
241727

Qy      646 CTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCT 705
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241726 CTTCTACTACAAAATGGTAGTCTTCTTAAAGAAGAGGAGCCAGCAGCAGGCAACTGTGCT
241667

Qy      706 GCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTT 765
      | ||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241666 ATCGCTGAACAAACCTCTGCGCCTGGTGGTCCTGGCAGTGGTGATCTTCTCTGTACTCTT
241607

Qy      766 CACACCCTATCATATCATGCGCAATTT 792
      |||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      241606 TACACCTTACCATATCATGCGCAATGT 241580

```

# RESULT 8

AF247785

LOCUS AF247785 1325 bp mRNA linear PRI 26-MAR-2002

DEFINITION Homo sapiens P2Y purinoceptor 1 mRNA, complete cds.

ACCESSION AF247785

VERSION AF247785.1 GI:19716154

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1325)

AUTHORS Zhang,W., Li,N., Wan,T. and Cao,X.

TITLE Human P2Y purinoceptor 1

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1325)

AUTHORS Zhang,W., Li,N., Wan,T. and Cao,X.

TITLE Direct Submission

JOURNAL Submitted (21-MAR-2000) Department of Immunology, Second Military  
Medical University & Shanghai Brilliance Biotechnology Institute,  
800 Xiangyin Rd., Shanghai 200433, P.R. China

FEATURES Location/Qualifiers

source

1. .1325

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

CDS

69. .1073

/codon\_start=1

/product="P2Y purinoceptor 1"

/protein\_id="AAL95690.1"

/db\_xref="GI:19716155"

/translation="MLGIMAWNATCKNWLAEEAALEKYYLSIFYGIEFVVGVLGNTIV  
VYGYIFSLKNWNSSNIYLFNLSVSDLAFLCTLPMLIRSYANGNWIYGDVLCISNRYVL



```

      |||| |||| || || || || |   | |||| || |||| || || |||||
Db      736 CTGCTCTGCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 795

Qy      759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818
      | || || ||||| ||||| ||| ||||| ||||| ||||| |||||
Db      796 TGCTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 855

Qy      819 G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC 875
      |  || |  || || || |  ||||| || | |||| | ||||| |||||
Db      856 GGAAGCAGTATCAGTGCACCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTT 915

Qy      876 TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACA 935
      ||||| ||||| ||||| || || ||||| || || ||||| || || ||
Db      916 TGGCCTTTCTGAACAGTGTCAACCTGTCTTCTATTTCTTTGGGAGATCACTTCA 975

Qy      936 GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT 995
      | || ||||| || | | | |||| | ||||| ||||| ||||| || || ||
Db      976 GGGACATGCTGATGAATCAACTGAGACACAACCTCAAATCCCTTACATCCTTTAGCAGAT 1035

Qy      996 GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG 1048
      | ||| || | |  |||| |||| |||| |||| |||| |||||
Db      1036 GGGCTCATGAACCTCTACTTTTCATTTCAGAGAAAAGTGAGGGGCTTGTGAAACAG 1089

```

# RESULT 9

AX549281

LOCUS AX549281 1380 bp DNA linear PAT 26-NOV-2002

DEFINITION Sequence 566 from Patent WO02061087.

ACCESSION AX549281

VERSION AX549281.1 GI:25813951

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Burmer,G.C., Roush,C.L. and Brown,J.P.

TITLE Antigenic peptides, such as for G protein-coupled receptors  
(GPCRs), antibodies thereto, and systems for identifying such  
antigenic peptides

JOURNAL Patent: WO 02061087-A 566 08-AUG-2002;  
Lifespan Biosciences, Inc. (US)

FEATURES

Location/Qualifiers  
source 1. .1380  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 383 a 294 c 274 g 429 t

ORIGIN

Query Match 38.4%; Score 592.4; DB 6; Length 1380;

Best Local Similarity 75.3%; Pred. No. 1.7e-121;

Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

Qy 39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98

Db 50 GGATCATGGCATGGAATGCAACTTGCAAAAAGTGGCTGGCAGCAGAGGCTGCCCTGGAAA 109

Qy 99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158  
 ||||| || ||||| || ||||| || | ||| | ||||| || ||  
 Db 110 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCCTGGAAATACCA 169  
 Qy 159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218  
 ||| || | ||||| ||||| ||||| ||||| ||||| ||||| ||  
 Db 170 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 229  
 Qy 219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278  
 ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 230 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCATGCTGATAAGGAGTT 289  
 Qy 279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338  
 ||||| || ||| ||||| || ||||| || ||||| ||||| ||||| |||||  
 Db 290 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 349  
 Qy 339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398  
 | ||||| ||||| ||||| ||||| ||||| || ||||| || ||||| ||  
 Db 350 ATGCCAACCTCTATACCAGCATCTCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 409  
 Qy 399 TCATGAAGTACCCTTTCCGAGAACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458  
 | || ||||| ||||| ||||| ||||| ||||| || ||||| ||||| |||||  
 Db 410 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 469  
 Qy 459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCAATCAATT 518  
 | |||| | |||| ||||| ||||| ||||| || ||||| || |||||  
 Db 470 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 529  
 Qy 519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAAC 578  
 |||| || || | |||| | |||| | || | ||||| ||||| ||||| |  
 Db 530 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGAAGTTCTGGAGACCCCAACT 589  
 Qy 579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638  
 |||| ||||| ||||| | || || || ||||| ||||| ||||| || |||||  
 Db 590 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTTGTGA 649  
 Qy 639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698  
 |||| |||| || ||||| || |||| ||||| ||||| |||| ||||  
 Db 650 TGTGTTTCTTTTATTACAAGATTGCTCTCTCCTAAAGCAGAGGAATAGGCAGGTTGCTA 709  
 Qy 699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758  
 |||| |||| || || || || || || |||| || |||| || || |||||  
 Db 710 CTGCTCTGCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 769  
 Qy 759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818  
 | || || ||||| ||||| ||||| || ||||| ||||| ||||| |||||  
 Db 770 TGCTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 829  
 Qy 819 G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC 875  
 | || || | || || || || |||| || || |||| ||||| |||||  
 Db 830 GGAAGCAGTATCAGTGCACCTCAGGTCGTCACTCAACTCCTTTTACATTGTGACACGGCCTT 889  
 Qy 876 TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCTCATGGGAGACCATTACA 935  
 ||||| ||||| ||||| || ||||| || || ||||| || || ||  
 Db 890 TGGCCTTTCTGAACAGTGTCATCAACCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA 949





Query Match 38.4%; Score 592.4; DB 9; Length 1380;  
Best Local Similarity 75.3%; Pred. No. 1.7e-121;  
Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      50 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 109

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     110 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 169

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     170 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 229

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     230 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCATGCTGATAAGGAGTT 289

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     290 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 349

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     350 ATGCCAACCTCTATACCAGCATTCTCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 409

Qy     399 TCATGAAGTACCCTTTCCGAGAACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     410 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 469

Qy     459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCACCCATGCTCACTTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     470 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 529

Qy     519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCCTGGAAACCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     530 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCCTGGAGACCCCACT 589

Qy     579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     590 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 649

Qy     639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     650 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCCTAAAGCAGAGGAATAGGCAGGTTGCTA 709

Qy     699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     710 CTGCTCTGCCCCTTGAAAAGCCTCTCAACTGGTCATCATGGCAGTGGTAATCTTCTCTG 769

Qy     759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     770 TGCTTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 829
```

Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	830	GGAAGCAGTATCAGTGCCTCAGGTCGTCACTCAACTCCTTTTACATTGTGACACGGCCTT	889
Qy	876	TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACA	935
Db	890	TGGCCTTTCTGAACAGTGTCAACCTGTCTTCTATTTCTTTTGGGAGATCACTTCA	949
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	950	GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT	1009
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG	1048
Db	1010	GGGCTCATGAACTCCTACTTTTCATTTCAGAGAAAAGTGAGGGGCTTGTGAAACAG	1063



Db 527 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAAATC 586  
 Qy 519 CTGTCCCAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCTGGAAACCCTGAAC 578  
 |||| || || | |||| | || | || | ||||| |||| |  
 Db 587 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCACT 646  
 Qy 579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTGGGCTTCCTAATTCCTCTCTCTGTGA 638  
 |||| ||||| |||| | || || || ||||| |||| | |||||  
 Db 647 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 706  
 Qy 639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGCCAGCAGCAAGCAA 698  
 |||| |||| || ||||| | |||| |||| |||| |||| || ||  
 Db 707 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 766  
 Qy 699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTA 758  
 |||| |||| || || || || | |||| || |||| || |||||  
 Db 767 CTGCTCTGCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 826  
 Qy 759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818  
 | || || ||||| ||||| || ||||| ||||| ||||| ||||  
 Db 827 TGCTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 886  
 Qy 819 G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC 875  
 | || | || || || | |||| || | |||| |||||  
 Db 887 GGAAGCAGTATCAGTGCCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTT 946  
 Qy 876 TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCCTCATGGGAGACCATTACA 935  
 ||||| ||||| |||| || ||||| || || ||||| || ||  
 Db 947 TGGCCTTTCTGAACAGTGTCTCAACCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA 1006  
 Qy 936 GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT 995  
 | || ||||| || || | |||| ||||| ||||| || || ||  
 Db 1007 GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT 1066  
 Qy 996 GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG 1048  
 | ||| || | | |||| |||| |||| |||| |||||  
 Db 1067 GGGCTCATGAACCTCTACTTTCATTCAGAGAAAAGTGAGGGGCTTGTGAAACAG 1120

# RESULT 12

AX342665

LOCUS AX342665 1542 bp DNA linear PAT 12-JAN-2002

DEFINITION Sequence 20 from Patent WO0198351.

ACCESSION AX342665

VERSION AX342665.1 GI:18152045

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Lal,P., Baughn,M.R., Hafalia,A.J., Nguyen,D.B., Gandhi,A.R., Kallick,D.A., Griffin,J.A., Yue,H., Khan,F.A., Patterson,C., Lu,D.A., Tribouley,C.M., Lu,Y., Walia,N.K., Graul,R., Yao,M.G., Yang,J., Ramkumar,J., Au-Young,J., Hernandez,R., Walsh,R.T. and Borowsky,M.L.

JOURNAL Patent: WO 0198351-A 20 27-DEC-2001;

Incyte Genomics, Inc. (US)  
 FEATURES Location/Qualifiers  
     source 1. .1542  
             /organism="Homo sapiens"  
             /mol\_type="genomic DNA"  
             /db\_xref="taxon:9606"  
             /note="Incyte ID No: 3485895CB1"  
 BASE COUNT 428 a 327 c 315 g 472 t  
 ORIGIN

Query Match 38.4%; Score 592.4; DB 6; Length 1542;  
 Best Local Similarity 75.3%; Pred. No. 1.7e-121;  
 Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

```

Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | ||||| ||| | ||| | || ||| ||||| ||||| | || | |
Db      205 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 264

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA 158
      ||||| || ||||| || ||||| || ||| | ||||| || ||
Db      265 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 324

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      ||| || | ||||| ||||| ||||| ||||| ||||| ||| | ||||| |
Db     325 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 384

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     385 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 444

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      ||||| || ||| ||||| || ||||| ||||| ||||| ||||| |||||
Db     445 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 504

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGC 398
      | ||||| ||||| ||||| ||||| ||||| || ||||| || ||||| ||
Db     505 ATGCCAACCTCTATACCAGCATTCTCTTCTCACTTTTATCAGCATAGATCGATACTTGA 564

Qy     399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     565 TAATTAAGTATCCTTTCCGAGAACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 624

Qy     459 CGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATT 518
      | ||| | ||| ||||| ||||| | ||||| || | | ||| |||
Db     625 CCTTGGCCATTTGGGTTTATAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 684

Qy     519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAAC 578
      |||| || || | |||| | ||| | || | ||||| ||||| |||| |
Db     685 CTGTTATAACTGACAAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT 744

Qy     579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      |||| ||||| ||||| | || || || ||||| ||||| ||||| |||||
Db     745 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 804

Qy     639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698
      |||| |||| | ||||| | |||| |||| | |||| | ||| || |
Db     805 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 864

```



Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 90343)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
REFERENCE 3 (bases 1 to 90343)  
AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (09-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
COMMENT INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence



continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:

<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .90343 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-3F11"
repeat_region	991. .1106 /rpt_family="MER45B"
repeat_region	complement(1314. .1627) /rpt_family="AluSx"
repeat_region	complement(2137. .2430) /rpt_family="AluY"
repeat_region	complement(2568. .2741) /rpt_family="L1M4"
repeat_region	complement(2742. .3047) /rpt_family="AluSx"
repeat_region	complement(3048. .3165) /rpt_family="L1M4"
repeat_region	4735. .4865 /rpt_family="FLAM_C"
repeat_region	5657. .5762 /rpt_family="L1MC/D"
repeat_region	5906. .6237 /rpt_family="LTR21B"
repeat_region	6289. .6773 /rpt_family="HERVFH21"
repeat_region	complement(8725. .9597) /rpt_family="MER11D"
STS	12399. .12689 /standard_name="136046"
repeat_region	13774. .13816 /rpt_family="Alu"
repeat_region	13817. .13874 /rpt_family="(TA)n"
repeat_region	complement(15157. .15633) /rpt_family="L2"
repeat_region	15706. .15747 /rpt_family="AT_rich"

```

repeat_region 16025. .16235
                /rpt_family="MIR"
repeat_region 16560. .16682
                /rpt_family="L2"
repeat_region complement(16710. .17265)
                /rpt_family="LTR49"
repeat_region 18077. .18368
                /rpt_family="AluSx"
repeat_region complement(18376. .18471)
                /rpt_family="L2"
repeat_region complement(18486. .18859)
                /rpt_family="MER57B"
repeat_region complement(20618. .20922)
                /rpt_family="AluSx"
repeat_region 21337. .21363
                /rpt_family="AT_rich"
repeat_region 22155. .22561
                /rpt_family="L1M4"
repeat_region complement(22608. .22659)
                /rpt_family="L1M4"
repeat_region 22685. .23013
                /rpt_family="L1MB8"
repeat_region 23103. .23399
                /rpt_family="AluSg"
repeat_region 23500. .23973
                /rpt_family="L1ME3A"
repeat_region complement(24027. .24305)
                /rpt_family="L1MB1"
repeat_region 24304. .24655
                /rpt_family="L1ME3A"
repeat_region 24656. .24697
                /rpt_family="MADE1"
repeat_region 25203. .25518
                /rpt_family="AluJo"
repeat_region 25783. .25817
                /rpt_family="(TAA)n"
repeat_region 26187. .26211
                /rpt_family="AT_rich"
repeat_region 27014. .27030
                /rpt_family="AT_rich"
repeat_region complement(27031. .27316)
                /rpt_family="AluSx"
repeat_region 27317. .27328
                /rpt_family="AT_rich"
repeat_region 27574. .27615
                /rpt_family="(TAGA)n"
STS            28062. .28166
                /standard_name="24707"
STS            28199. .28382
                /standard_name="13170"
repeat_region complement(29079. .29167)
                /rpt_family="MLT1J"
repeat_region 29168. .29532
                /rpt_family="THE1B"
repeat_region complement(29533. .29552)
                /rpt_family="MLT1J"
repeat_region 29807. .30387

```

Query Match 38.3%; Score 590.2; DB 9; Length 90343;  
Best Local Similarity 75.5%; Pred. No. 6e-121;  
Matches 760; Conservative 0; Mismatches 243; Indels 4; Gaps 2;

```
Qy      46 GGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTA 105
      ||||  ||||  | ||||  | || ||| ||||| ||||| || || | |||||
Db      80664 GGCATGGAATGCAACTTGCAAAAAGTGGCTGGCAGCAGAGGCTGCCCTGGAAAAGTACTA 80723

Qy     106 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGT 165
      ||| ||  |||||  || |||||  || | ||| | |||| |||  || ||| ||
Db     80724 CCTTTCCATTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCATTGTTGT 80783

Qy     166 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCT 225
      | |||||  |||||  ||||| ||||| ||||| ||||| ||| | ||||| |||||
Db     80784 TTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCTTAAACCT 80843

Qy     226 TTCCATCTCTGACTTTGCTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA 285
      ||  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     80844 CTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCATGCTGATAAGGAGTTATGCCAA 80903

Qy     286 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA 345
      ||  ||  |||  |||||  || |||||  ||||| ||||| ||||| ||||| |||||
Db     80904 TGGAAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTCATGCCAA 80963

Qy     346 CCTCTACACCAGCATCCTCTTCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAA 405
      |||||  |||||  ||||| ||||| ||||| || |||||  || |||||  || || ||
Db     80964 CCTCTATACCAGCATTCTCTTCTCACTTTTATCAGCATAGATCGATACTTGATAATTAA 81023

Qy     406 GTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGTGCGC 465
      |||  ||||| ||||| ||||| ||||| || |||||  ||||| ||||| |||||
Db     81024 GTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGTATTTTAATCTCCTTGGC 81083

Qy     466 TGTCTGGGCCTTAGTGACCTTAGAAGTTCACCCATGCTCACTTTCATCAATTCTGTCCC 525
      | ||||  ||||  |||||  | |||||  || | || ||| ||| ||||
Db     81084 CATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATCCTGTTAT 81143

Qy     526 AAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTCTGGAAACCCTGAACACAATCT 585
      ||  ||  | ||||  | |||  | || | ||| ||||| ||||  | |||| ||
Db     81144 AACTGACAATGGCACCACCTGTAATGATTTTGAAGTCTGGAGACCCCACTACAACCT 81203

Qy     586 CATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTT 645
      ||||| ||||  | || || ||  |||||  ||||  ||||| || ||||| || ||
Db     81204 CATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGATGTGTTT 81263

Qy     646 CTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCT 705
      ||| ||  |||||  |  ||||  ||||  ||||  ||||  |||  || |||| ||
Db     81264 CTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTACTGCTCT 81323

Qy     706 GCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTT 765
      ||| ||  || || || || |  | ||||  || ||||  || || ||||| || || ||
Db     81324 GCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTGTGCTTTT 81383

Qy     766 CACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTG---GCC 822
      ||||| ||||  ||||| ||  ||||| ||||| ||||| ||||| ||||| ||
Db     81384 TACACCCTATCACGTATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTTGGAAGCA 81443
```

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsf., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,

Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,  
 Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,  
 Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,  
 Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,  
 Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,  
 Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,  
 Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,  
 Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,  
 Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,  
 Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,  
 Warren,R., Washington,C., Watlington,S., Williams,G.,  
 Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,  
 Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and  
 Gibbs,R.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 132745)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAY-2000) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 3 (bases 1 to 132745)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 4 (bases 1 to 132745)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 5 (bases 1 to 132745)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 6 (bases 1 to 132745)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 REFERENCE 7 (bases 1 to 132745)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Mar 28, 2002 this sequence version replaced gi:19718616.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .132745 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="3" /clone="RP11-64D22"
misc_feature	1. .2005 /note="overlaps bases 170209. .172213 of clone AC069067" /function="clone overlap"
STS	30. .130 /standard_name="74493"
repeat_region	complement(522. .1015) /rpt_family="MLT1D"
repeat_region	complement(2452. .2697) /rpt_family="L1MA5A"
repeat_region	complement(3200. .3578) /rpt_family="MLT1B"
repeat_region	3600. .3749 /rpt_family="(TA)n"
repeat_region	4391. .4411 /rpt_family="AT_rich"
repeat_region	4909. .4960

```

repeat_region    /rpt_family="AT_rich"
                  complement(5657. .6403)
                  /rpt_family="L1PA13"
repeat_region    6404. .7799
                  /rpt_family="L1PA13"
STS              8045. .8318
                  /standard_name="183647"
repeat_region    complement(8708. .9282)
                  /rpt_family="L1MD3"
repeat_region    complement(9287. .9357)
                  /rpt_family="MLT1F1"
repeat_region    complement(9359. .9460)
                  /rpt_family="L1MC3"
repeat_region    complement(9587. .9880)
                  /rpt_family="AluSg"
repeat_region    complement(10203. .10450)
                  /rpt_family="L1MC4"
repeat_region    11468. .11699
                  /rpt_family="L1M4"
repeat_region    11717. .11886
                  /rpt_family="MLT2CB"
repeat_region    11908. .11982
                  /rpt_family="Tigger3(Golem)"
repeat_region    12020. .12246
                  /rpt_family="THE1C"
repeat_region    12263. .12562
                  /rpt_family="AluSx"
repeat_region    13326. .13346
                  /rpt_family="AT_rich"
repeat_region    complement(13934. .14245)
                  /rpt_family="AluSx"
repeat_region    complement(14256. .14568)
                  /rpt_family="AluJo"
repeat_region    14618. .14746
                  /rpt_family="MER8"
repeat_region    14825. .14849
                  /rpt_family="AT_rich"
repeat_region    14865. .14906
                  /rpt_family="AT_rich"
repeat_region    15465. .15739
                  /rpt_family="L2"
repeat_region    16579. .16756
                  /rpt_family="(TTATA)n"
repeat_region    complement(16757. .17074)
                  /rpt_family="L2"
repeat_region    17621. .17660
                  /rpt_family="(CAAAA)n"
repeat_region    18544. .18725

```

```

Query Match          38.3%;  Score 590.2;  DB 9;  Length 132745;
Best Local Similarity 75.5%;  Pred. No. 6.1e-121;
Matches 760;  Conservative 0;  Mismatches 243;  Indels 4;  Gaps 2;

```

```

Qy          46  GGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTA 105
              ||||  ||||  | ||||  | ||  || ||||  ||||  ||||  ||||  ||||  ||||
Db    123065  GGCATGGAATGCAACTTGCAAAAAGTGGCTGGCAGCAGAGGCTGCCCTGGAAAAGTACTA
123124

```

Qy 106 CCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGT 165  
 ||| || ||||| || ||||| || | ||| | |||| ||| || ||| ||  
 Db 123125 CCTTTCCATTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCATTGTTGT  
 123184

Qy 166 GTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACCT 225  
 | ||||| ||||| ||||| ||||| ||||| ||| | |||| |||||  
 Db 123185 TTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTATCTCTTTAACCT  
 123244

Qy 226 TTCCATCTCTGACTTTGCTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAA 285  
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Db 123245 CTCTGTCTCTGACTTAGCTTTCTGTGCACCCTCCCATGCTGATAAGGAGTTATGCCAA  
 123304

Qy 286 TGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAA 345  
 || || ||| ||||| || ||||| ||||| ||||| ||||| |||||  
 Db 123305 TGGAAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTCATGCCAA  
 123364

Qy 346 CCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAA 405  
 ||||| ||||| ||||| ||||| || ||||| || ||||| || |||||  
 Db 123365 CCTCTATACCAGCATTCTCTTTCTCACTTTTATCAGCATAGATCGATACTTGATAATTAA  
 123424

Qy 406 GTACCCTTTCCGAGAACACTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGC 465  
 || ||||| ||||| ||||| ||||| || ||||| || ||||| |||||  
 Db 123425 GTATCCTTTCCGAGAACACTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCTCCTTGGC  
 123484

Qy 466 TGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCC 525  
 | |||| ||||| ||||| | ||||| || | | || |||||  
 Db 123485 CATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACCTCCCTTATAAATCCTGTTAT  
 123544

Qy 526 AAAAGAAGAGGGCAGTAAGTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCT 585  
 || || | |||| | ||| | || | ||||| |||| | |||| ||  
 Db 123545 AACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCACTACAACCT  
 123604

Qy 586 CATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTT 645  
 ||||| || | || || || ||||| |||| ||||| | ||||| ||  
 Db 123605 CATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTTGTGATGTGTTT  
 123664

Qy 646 CTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCT 705  
 ||| || ||||| | |||| |||| |||| |||| || ||||| ||  
 Db 123665 CTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTACTGCTCT  
 123724

Qy 706 GCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTT 765  
 ||| || || || || | |||| || |||| || || ||||| || || ||  
 Db 123725 GCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGAATCTTCTCTGTGCTTTT  
 123784

Qy 766 CACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTG--GCC 822





Db	752	TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT	811
Qy	219	TTAACCTTTCCATCTCTGACTTTGCTTTCCCTGTGCACCCTTCCCATCCTGATAAAGAGTT	278
Db	812	TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT	871
Qy	279	ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC	338
Db	872	ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC	931
Qy	339	ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC	398
Db	932	ATGCCAACCTCTATACCAGCATTCCTTTCTCACTTTTATCAGCATAGATCGATACTTGA	991
Qy	399	TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT	458
Db	992	TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT	1051
Qy	459	CGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT	518
Db	1052	CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC	1111
Qy	519	CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCTGAAC	578
Db	1112	CTGTTATAACTGACAATGGCACACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT	1171
Qy	579	ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA	638
Db	1172	ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA	1231
Qy	639	TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA	698
Db	1232	TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA	1291
Qy	699	CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA	758
Db	1292	CTGCTCTGCCCCTTGAAAAGCCTCTCAACTGGTCATCATGGCAGTGGTAATCTTCTCTG	1351
Qy	759	TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT	818
Db	1352	TGCTTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT	1411
Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	1412	GGAAGCAGTATCAGTGCACCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGGCTT	1471
Qy	876	TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCCTCATGGGAGACCATTACA	935
Db	1472	TGGGCTTTCTGAACAGTGTCATCAACCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA	1531
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	1532	GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT	1591
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG	1048

Db 1592 GGGCTCATGAACTCCTACTTTCATTCAGAGAAAAGTGAGGGGCTTGTGAAACAG 1645

Search completed: December 14, 2003, 15:00:20  
Job time : 5884 secs

OM nucleic - nucleic search, using sw model

Run on: December 14, 2003, 13:14:29 ; Search time 3520 Seconds  
(without alignments)  
10653.922 Million cell updates/sec

Title: US-09-891-138A-1  
Perfect score: 1543  
Sequence: 1 gctcctggcagagttttctg.....tgccataaataaatcaatata 1543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	%		Query		DB		ID	Description
	No.	Score	Match	Length				
	1	1484.8	96.2	1585	11	AK080866		AK080866 Mus muscu
	2	560	36.3	683	10	BB323771		BB323771 BB323771
	3	516.4	33.5	556	4	BX527630		Bx527630 RZPD Mus
	4	495.8	32.1	520	9	AI663305		AI663305 uk27c10.y
	5	455	29.5	469	10	BB744515		BB744515 BB744515
	6	438	28.4	458	10	BB746222		BB746222 BB746222
	7	414	26.8	428	10	BB738743		BB738743 BB738743
	8	403.8	26.2	422	10	BB847918		BB847918 BB847918
	9	388.4	25.2	420	10	BB864882		BB864882 BB864882
	10	384.8	24.9	426	10	BB778587		BB778587 BB778587
	11	380.4	24.7	396	10	BB739482		BB739482 BB739482
c	12	363.8	23.6	367	9	AI649254		AI649254 uk27c10.x
	13	357.6	23.2	636	10	BB645274		BB645274 BB645274
	14	354.2	23.0	416	10	BB846608		BB846608 BB846608
	15	350.6	22.7	408	13	BY368584		BY368584 BY368584
	16	346.6	22.5	868	14	CD246161		CD246161 AGENCOURT
	17	309.8	20.1	327	10	BB220946		BB220946 BB220946
	18	296	19.2	377	10	BG402029		BG402029 602465748
	19	294.6	19.1	333	10	BB254869		BB254869 BB254869
	20	283.2	18.4	323	10	BB220888		BB220888 BB220888
	21	279	18.1	316	10	BB225749		BB225749 BB225749
	22	276.2	17.9	501	13	BX281458		BX281458 BX281458
	23	275.2	17.8	326	10	BB500452		BB500452 BB500452
	24	269.4	17.5	770	13	BU141159		BU141159 603137524
	25	261.4	16.9	285	10	BB327439		BB327439 BB327439
c	26	255.8	16.6	633	29	AG083174		AG083174 Pan trogl
	27	253.4	16.4	314	10	BB498575		BB498575 BB498575
	28	241	15.6	310	10	BB221521		BB221521 BB221521
	29	218.4	14.2	805	13	BU291924		BU291924 603606116
	30	216.8	14.1	366	13	BY005778		BY005778 BY005778
	31	215.4	14.0	282	10	BB215653		BB215653 BB215653
	32	214	13.9	312	10	BB498898		BB498898 BB498898
	33	202	13.1	1101	29	CNS04VV3		AL309576 Tetraodon
	34	199.8	12.9	226	10	BB213317		BB213317 BB213317
c	35	192.8	12.5	956	29	CNS028Y4		AL186565 Tetraodon
	36	189.6	12.3	655	13	BU373390		BU373390 603811071
c	37	181	11.7	294	9	AW112068		AW112068 MC15648 m
c	38	158.8	10.3	639	9	AW612141		AW612141 hg94h07.x
	39	151.2	9.8	657	13	BU352057		BU352057 603527490
	40	151	9.8	1026	29	CNS051MY		AL317059 Tetraodon
c	41	144.2	9.3	589	10	BF196066		BF196066 hr81f02.x
	42	141	9.1	1022	29	CNS04W90		AL310077 Tetraodon
c	43	139.8	9.1	582	10	BE221739		BE221739 hr58c09.x
	44	137.4	8.9	139	9	AI021184		AI021184 ub02f12.r
	45	135.4	8.8	541	13	BU294029		BU294029 604162475

## ALIGNMENTS

## RESULT 1

AK080866

LOCUS AK080866 1585 bp mRNA linear HTC 05-DEC-2002

DEFINITION Mus musculus 4 days neonate male adipose cDNA, RIKEN full-length / enriched library, clone:B430012021 product:G-PROTEIN COUPLED RECEPTOR GPR91, full insert sequence.

ACCESSION AK080866

VERSION AK080866.1 GI:26099527

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE 1

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

## REFERENCE 2

AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

## REFERENCE 3

AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagacka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

## REFERENCE 4

AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,

Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyc-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.

TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409 (6821), 685-690 (2001)  
MEDLINE 21085660  
PUBMED 11217851  
REFERENCE 5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE 6 (bases 1 to 1585)  
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission  
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.go.jp/  
URL:http://fantom.gsc.riken.go.jp/.

FEATURES  
source Location/Qualifiers  
1. .1585  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:B430012021"  
/db\_xref="taxon:10090"  
/clone="B430012021"  
/sex="male"  
/tissue\_type="adipose"

```

                                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                                /dev_stage="4 days neonate"
misc_feature      69. .1025
                                /note="G-PROTEIN COUPLED RECEPTOR GPR91 (SPTR|Q99MT6,
                                evidence: FASTY, 94.3%ID, 100%length, match=954)
                                putative"
polyA_signal      1558. .1563
                                /note="putative"
polyA_site        1585
                                /note="putative"
BASE COUNT      450 a      351 c      305 g      477 t      2 others
ORIGIN

```

```

Query Match          96.2%;  Score 1484.8;  DB 11;  Length 1585;
Best Local Similarity 98.4%;  Pred. No. 1.5e-296;
Matches 1521;  Conservative 0;  Mismatches 22;  Indels 3;  Gaps 2;

```

```

Qy      1 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60
      |||
Db      26 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 35
      |||

Qy      61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120
      |||
Db      86 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGGATTTTA 145
      |||

Qy      121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTT 180
      |||
Db      146 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGGGGTGTTCGGGTACCTGTT 205
      |||

Qy      181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 240
      |||
Db      206 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 265
      |||

Qy      241 TGCTTTCTGTGCACCCTTCCCATCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300
      |||
Db      266 TGCTTTCTGTTGCACCCTTCCCATCTGATAAAGAGTTTGGCAATGATAAGGGGACCTA 325
      |||

Qy      301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360
      |||
Db      326 TGGAGATGTTCTTTGGATAAGCAACCGATATGGGCTTAACACCAACCTTTAAACCAGCAT 385
      |||

Qy      361 CCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAA--GTACCCTTTCCGA 418
      |||
Db      386 CTFTTTCTTCATTTTCATTAGCATGGACCGATATCTGCTCATGAAAGTACCCTTTTCCGA 445
      |||

Qy      419 GAACAC-TTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTT 477
      |||
Db      446 GAACACTTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTT 505
      |||

Qy      478 AGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATTCTGTCCCAAAAGAAGAGG 537
      |||
Db      506 AGTGACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATTCTGTCCCAAAAGAAGAGG 565
      |||

Qy      538 CAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCT 597
      |||
Db      566 CAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCT 625
      |||

```



Qy 598 CTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAA 657  
 |||  
 Db 626 CTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAA 685  
 |||  
 Qy 658 GATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAA 717  
 |||  
 Db 686 GATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAA 745  
 |||  
 Qy 718 ACCCCAACGCCTGGTGGTCTTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCA 777  
 |||  
 Db 746 ACCCCAACGCCTGGTGGTCTTGGCAGTTGTGATCTTCTCTATACTCTTCACACCCTATCA 805  
 |||  
 Qy 778 TATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACA 837  
 |||  
 Db 806 TATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACA 865  
 |||  
 Qy 838 GAAGGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCAT 897  
 |||  
 Db 866 GAAGGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCAT 925  
 |||  
 Qy 898 CAATCCCATCTTCTACTTCCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTT 957  
 |||  
 Db 926 CAATCCCATCTTCTACTTCCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTT 985  
 |||  
 Qy 958 CAGACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTC 1017  
 |||  
 Db 986 CAGACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTC 1045  
 |||  
 Qy 1018 ACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTTAACTAAGTAA 1077  
 |||  
 Db 1046 ACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTTAACTAAGTAA 1105  
 |||  
 Qy 1078 ACCACCATTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCAGGGCTGGAGTACAAG 1137  
 |||  
 Db 1106 ACCACCATTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCAGGGCTGGAGTACAAG 1165  
 |||  
 Qy 1138 CTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTA 1197  
 |||  
 Db 1166 CTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTA 1225  
 |||  
 Qy 1198 TGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGA 1257  
 |||  
 Db 1226 TGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGA 1285  
 |||  
 Qy 1258 ACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAATTTAAGACCTCT 1317  
 |||  
 Db 1286 ACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAATTTAAGACCTCT 1345  
 |||  
 Qy 1318 TTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCAT 1377  
 |||  
 Db 1346 TTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCAT 1405  
 |||  
 Qy 1378 CATGGTTCAGGTCGATAAGCGTGTCTTCTGAAATAGTCTTATTTTATTCTTGTAATATTA 1437  
 |||  
 Db 1406 CATGGTTCAGGTCGATAAGCGTGTCTTCTGAAATAGTCTTATTTTATTCTTGTAATATTA 1465  
 |||  
 Qy 1438 AAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAAAT 1497  
 |||

Db 1466 ||||| AAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAAT 1525

Qy 1498 TATATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543

Db 1526 ||||| TATATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1571

# RESULT 2

BB323771

LOCUS BB323771 683 bp mRNA linear EST 31-AUG-2001

DEFINITION BB323771 RIKEN full-length enriched, 4 days neonate male adipose  
Mus musculus cDNA clone B430012021 3', mRNA sequence.

ACCESSION BB323771

VERSION BB323771.2 GI:15411432

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 683)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished

COMMENT On Jul 11, 2000 this sequence version replaced gi:9032085.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .  
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

```

FEATURES             Location/Qualifiers
     source            1..683
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /db_xref="taxon:10090"
                        /clone="B430012021"
                        /sex="male"
                        /tissue_type="adipose"
                        /dev_stage="4 days neonate"
                        /lab_host="DH10B"
                        /clone_lib="RIKEN full-length enriched, 4 days neonate
                        male adipose"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTVN 3']; cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. cDNA went through one round of normalization
                        to Rot = 10.0 and subtraction to Rot = 229.0. Second
                        strand cDNA was prepared with the primer adapter of
                        sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC
                        3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                        modified pBluescript KS(+) after bulk excision from Lambda
                        FLC I."

```

```

BASE COUNT      226 a    125 c    117 g    215 t
ORIGIN

```

```

Query Match      36.3%; Score 560; DB 10; Length 683;
Best Local Similarity 98.7%; Pred. No. 2.5e-105;
Matches: 596; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```

```

Qy      943 GCTGATTAGTAAGTTCAGAC-AATACTTCAAG-TCCCTTACATCCTTC-AGGACATGAGC 999
          |||
Db      66 GCTGATTAGTAAGTTCAGCCAAATACTTCAAGTTCCTTACATCCTTCAAGGACATAAGT 125

Qy     1000 TGCTGGATGCAGGTCTTCACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGT 1059
          |||
Db     125 TGCTGGATGCAGGTTTTCACTCAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGT 185

Qy     1060 TGAGTTTAACTAAGTAAACCACCATTCTAGGCTTTAGCTTTCCACCATCCTCCAACCC 1119

```

```

Db      185  |||||TGAGTTTAAATTAAGTAAACCACCATTCTAGGCTTTAGCTTTCCACCATCCTCCAACCC 245
Qy      1120 CCAGGGCTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTT 1179
Db      246  |||||CCAGGGCTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTT 305
Qy      1180 TAGGTTTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTA 1239
Db      306  |||||TAGGTTTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTA 365
Qy      1240 AGAACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTT 1299
Db      366  |||||AGAACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTT 425
Qy      1300 GGAAATTTTAAGACCTCTTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAA 1359
Db      426  |||||GGAAATTTTAAGACCTCTTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAA 485
Qy      1360 ATGCTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATT 1419
Db      486  |||||ATGCTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATT 545
Qy      1420 TTTATTCTTGTAATATTTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTA 1479
Db      546  |||||TTTATTCTTGTAATATTTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTA 605
Qy      1480 GATTTTCTATTTGAAAATTATATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAA 1539
Db      606  |||||GATTTTCTATTTGAAAATTATATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAA 565
Qy      1540 TATA 1543
Db      666  |||||TATA 669

```

# RESULT 3

BX527630

ID BX527630 standard; RNA; EST; 556 BP.

XX

AC BX527630;

XX

SV BX527630.1

XX

DT 27-MAY-2003 (Rel. 75, Created)

DT 27-MAY-2003 (Rel. 75, Last updated, Version 1)

XX

DE RZPD Mus musculus cDNA clone IMAGp998B194840 = IMAGE:1970226 5' EST.

XX

KW EST; expressed sequence tag.

XX

OS Mus musculus (house mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

XX

RN [1]

RP 1-556

RA Heil O., Ebert L., Neubert P., Peters M., Radelof U., Schneider D.,  
RA Korn B.;  
RT ;  
RL Submitted (28-MAY-2003) to the EMBL/GenBank/DDBJ databases.  
RL RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer  
RL Feld 580, D-69120 Heidelberg, Germany

XX

CC RZPD; IMAGp998B194840.  
CC RZPDLIB; I.M.A.G.E. cDNA Clone Collection;  
CC Mouse UnigeneSet - RZPD2 (RZPDLIB No.981)  
CC <http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981>  
CC Contact: Ina Rolfs  
CC RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
CC Heubnerweg 6, D-14059 Berlin, Germany  
CC Tel: +49 30 32639 101  
CC Fax: +49 30 32639 111  
CC [www.rzpd.de](http://www.rzpd.de)  
CC This clone is available royalty-free from RZPD;  
CC contact RZPD ([clone@rzpd.de](mailto:clone@rzpd.de)) for further information.  
CC Seq primer: sugF, Primer sequence: CTTCTGCTCTAAAAGCTGCG

XX

FH Key Location/Qualifiers

FH

FT source 1. .556  
FT /db\_xref="taxon:10090"  
FT /note="1st strand cDNA was primed with an oligo(dT) primer  
FT [ATGTGGCCTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was  
FT ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and  
FT cloned into distinct DraIII sites of the pME18S-FL3 vector  
FT (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used  
FT to isolate the cDNA insert. Size selection was performed to  
FT exclude fragments <1.5kb. Library constructed by Dr. Sumio  
FT Sugano (University of Tokyo Institute of Medical Science).  
FT Custom primers for sequencing: 5' end primer  
FT CTTCTGCTCTAAAAGCTGCG and 3' end primer  
FT CGACCTGCAGCTCGAGCACA. REFERENCES: Suzuki, Y., Yoshitomo,  
FT K., Maruyama, K., Suyama, A., and Sugano, S. Construction  
FT and characterization of a full length-enriched and a 5' end  
FT enriched cDNA library. Gene 200, 149-156, 1997. Sasaki, Z.,  
FT Suzuki, Y., Watanabe, M., Imai, J., Shibui, A., Yoshida,  
FT K., Hata. H., Yamaguchi, R., Tateyama, S., and Sugano, S.  
FT Construction of mouse full length-enriched cDNA libraries  
FT by oligo-capping. DNA Research, submitted."  
FT /organism="Mus musculus"  
FT /clone="IMAGp998B194840"  
FT /clone\_lib="Sugano mouse kidney mkia"  
FT /dev\_stage="adult"  
FT /lab\_host="DH10B"

XX

SQ Sequence 556 BP; 134 A; 137 C; 111 G; 173 T; 1 other;

Query Match 33.5%; Score 516.4; DB 4; Length 556;  
Best Local Similarity 99.8%; Pred. No. 2.5e-96;  
Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60  
|||||

Db 39 GCTCCTGGCAGAGTTTCTGTGTCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 98  
 Qy 61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120  
 Db 99 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 158  
 Qy 121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTT 180  
 Db 159 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTT 218  
 Qy 181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTAACCTTTCCATCTCTGACTT 240  
 Db 219 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTAACCTTTCCATCTCTGACTT 278  
 Qy 241 TGCTTTCCTGTGCACCCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300  
 Db 279 TGCTTTCCTGTGCACCCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 338  
 Qy 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360  
 Db 339 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 398  
 Qy 361 CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA 420  
 Db 399 CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA 458  
 Qy 421 AACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGCTGCGGCCTTAGT 480  
 Db 459 AACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGCTGCGGCCTTAGT 518  
 Qy 481 GACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATT 518  
 Db 519 GACCTTAGAAGTTCTACCCATGCTCACTTTTCATCAATT 556

#### RESULT 4

AI663305

LOCUS AI663305 520 bp mRNA linear EST 10-MAY-1999

DEFINITION uk27c10.y1 Sugano mouse kidney mkia Mus musculus cDNA clone  
 IMAGE:1970226 5' similar to SW:P2YR\_RAT P49651 P2Y PURINOCEPTOR 1  
 ;, mRNA sequence.

ACCESSION AI663305

VERSION AI663305.1 GI:4766888

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 520)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
 ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
 ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
 Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished

COMMENT Other\_ESTs: uk27c10.x1

Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:986966  
 Seq primer: custom primer used  
 High quality sequence stop: 490.

FEATURES                      Location/Qualifiers  
     source                    1. .520  
                               /organism="Mus musculus"  
                               /mol\_type="mRNA"  
                               /strain="C57BL"  
                               /db\_xref="taxon:10090"  
                               /clone="IMAGE:1970226"  
                               /sex="female"  
                               /dev\_stage="adult"  
                               /lab\_host="DH10B"  
                               /clone\_lib="Sugano mouse kidney mkia"  
                               /note="Organ: kidney; Vector: pME18S-FL3; Site\_1: DraIII  
                               (CACTGTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
                               was primed with an oligo(dT) primer  
                               [ATGTGGCCTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was  
                               ligated to a DraIII adaptor [TGTGGCCTACTGG]; digested  
                               and cloned into distinct DraIII sites of the pME18S-FL3  
                               vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should  
                               be used to isolate the cDNA insert. Size selection was  
                               performed to exclude fragments <1.5kb. Library  
                               constructed by Dr. Sumio Sugano (University of Tokyo  
                               Institute of Medical Science). Custom primers for  
                               sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end  
                               primer CGACCTGCAGCTCGAGCACA."

BASE COUNT        127 a     126 c     107 g     160 t  
 ORIGIN

Query Match                      32.1%;    Score 495.8;    DB 9;    Length 520;  
 Best Local Similarity    98.6%;    Pred. No. 4.6e-92;  
 Matches 500;    Conservative    0;    Mismatches    7;    Indels        0;    Gaps        0;

```

Qy      1 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60
        |||
Db      14 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCTGAATGGCACAGAATTTATC 73

Qy      61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120
        |||
Db      74 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 133

Qy     121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCTT 180
        |||
Db     134 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCTT 193

Qy     181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 240
        |||
Db     194 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 253
  
```

Qy 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300  
 |||||  
 Db 254 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 313  
 Qy 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360  
 |||||  
 Db 314 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 373  
 Qy 361 CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCGGAGA 420  
 |||||  
 Db 374 GCTCTTGCTCACTGTCTATTATCATGGACCGATATCTGCTCATGAAGTACCCTGTCCGAGA 433  
 Qy 421 ACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT 480  
 |||||  
 Db 434 ACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT 493  
 Qy 481 GACCTTAGAAGTTCTACCCATGCTCAC 507  
 |||||  
 Db 494 GACCTTAGAAGTTCTACCCATGCTCAC 520

# RESULT 5

BB744515

LOCUS BB744515 469 bp mRNA linear EST 16-OCT-2001

DEFINITION BB744515 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone F530003I24 3', mRNA sequence.

ACCESSION BB744515

VERSION BB744515.1 GI:16152351

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 469)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.



Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

```
FEATURES
    source                Location/Qualifiers
                        1. .469
                        /organism="Mus musculus"
                        /mol_type="mRNA"
                        /db_xref="taxon:10090"
                        /clone="F530003I24"
                        /sex="male"
                        /tissue_type="kidney"
                        /dev_stage="adult"
                        /lab_host="SOLR"
                        /clone_lib="RIKEN full-length enriched, adult male kidney"
                        /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. Second strand cDNA was prepared with the
                        primer adapter of sequence [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAACCCCCCCCCCCC 3'].
                        cDNA was cleaved with XhoI and SstI. "
```

BASE COUNT      160 a      75 c      80 g      154 t  
ORIGIN

Query Match                      29.5%;    Score 455;    DB 10;    Length 469;  
Best Local Similarity    100.0%;    Pred. No. 1.3e-83;  
Matches 455;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

```
QY            1089 TAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTGGGTCCACA 1148
             |||
Db            1 TAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTGGGTCCACA 60

QY            1149 TGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATA 1208
             |||
Db            61 TGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATA 120
```

Qy 1209 AGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACAAATATTGT 1268  
 |||  
 Db 121 AGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACAAATATTGT 180  
 Qy 1269 CAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTTCTATCAG 1328  
 |||  
 Db 181 CAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTTCTATCAG 240  
 Qy 1329 TGTAAAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGG 1388  
 |||  
 Db 241 TGTAAAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGG 300  
 Qy 1389 TCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATTCTTGTAATATTAAAATTTATGTG 1448  
 |||  
 Db 301 TCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATTCTTGTAATATTAAAATTTATGTG 360  
 Qy 1449 AAAAATGAATATAAATCAATGTACAACATTAGATTTTCTATTTGAAAATTATATTTCTTG 1508  
 |||  
 Db 361 AAAAATGAATATAAATCAATGTACAACATTAGATTTTCTATTTGAAAATTATATTTCTTG 420  
 Qy 1509 AAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543  
 |||  
 Db 421 AAAAAATAACTGCTGTGCCTAAATAAATCAATATA 455

# RESULT 6

BB746222

LOCUS BB746222 458 bp mRNA linear EST 15-OCT-2001

DEFINITION BB746222 RIKEN full-length enriched, adult male kidney *Mus musculus* cDNA clone F530013P03 3', mRNA sequence.

ACCESSION BB746222

VERSION BB746222.1 GI:16149159

KEYWORDS EST.

SOURCE *Mus musculus* (house mouse)

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; *Mus*.

REFERENCE 1 (bases 1 to 458)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
 e mouse tissues.

FEATURES                      Location/Qualifiers  
 source                      1. .458  
                               /organism="Mus musculus"  
                               /mol\_type="mRNA"  
                               /db\_xref="taxon:10090"  
                               /clone="F530013P03"  
                               /sex="male"  
                               /tissue\_type="kidney"  
                               /dev\_stage="adult"  
                               /lab\_host="SOLR"  
                               /clone\_lib="RIKEN full-length enriched, adult male kidney"  
                               /note="Site\_1: XhoI; Site\_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAACCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. "

BASE COUNT            150 a        75 c        82 g        151 t  
 ORIGIN

Query Match            28.4%;    Score 438;    DB 10;    Length 458;  
 Best Local Similarity   99.8%;    Pred. No. 4.1e-80;  
 Matches   449;    Conservative    0;    Mismatches    0;    Indels    1;    Gaps    1;

Qy            1058 GTTGAGTTTAACTAAGTAAACCACCATTCTAGGCTTTAGCTTTCCACCATCCTCCAAC 1117  
               |||||  
 Db            10 GTTGAGTTTAACTAAGTAAACCACCATTCTAGGCTTTAGCTTTCCACCATCCTCCAAC 69

Qy 1118 CCCCAGGGCTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGAT 1177  
 |||  
 Db 70 CCCCAGGGCTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGAT 129

Qy 1178 TTTAGGTTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACT 1237  
 |||  
 Db 130 TTTAGGTTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACT 189

Qy 1238 TAAGAACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATC 1297  
 |||  
 Db 190 TAAG-ACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATC 248

Qy 1298 TTGGAAATTTTAAGACCTCTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGC 1357  
 |||  
 Db 249 TTGGAAATTTTAAGACCTCTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGC 308

Qy 1358 AAATGCTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTA 1417  
 |||  
 Db 309 AAATGCTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTA 368

Qy 1418 TTTTATTCTTGTAATATTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACAT 1477  
 |||  
 Db 369 TTTTATTCTTGTAATATTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACAT 428

Qy 1478 TAGATTTTCTATTTGAAAATTATATTTCTT 1507  
 |||  
 Db 429 TAGATTTTCTATTTGAAAATTATATTTCTT 458

# RESULT 7

BB738743

LOCUS BB738743 428 bp mRNA linear EST 15-OCT-2001

DEFINITION BB738743 RIKEN full-length enriched, 6 days neonate spleen Mus musculus cDNA clone F430109C18 3', mRNA sequence.

ACCESSION BB738743

VERSION BB738743.1 GI:16141748

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 428)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

#### FEATURES

source

Location/Qualifiers

1. 428

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="F430109C18"

/tissue\_type="spleen"

/dev\_stage="6 days neonate"

/clone\_lib="RIKEN full-length enriched, 6 days neonate spleen"

BASE COUNT 153 a 59 c 72 g 144 t  
ORIGIN

Query Match 26.8%; Score 414; DB 10; Length 428;

Best Local Similarity 100.0%; Pred. No. 3.7e-75;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 AGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATAC 1189

Db 1 AGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATAC 60

Qy 1190 CCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAAC 1249

Db 61 CCAGAGTATGGAAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAAC 120

Qy 1250 AAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAATTTTA 1309

Db 121 AAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGGAAATTTTA 180

Qy 1310 AGACCTCTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATG 1369

```

Db      181  |||||
          AGACCTCTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATG 240

Qy      1370 CATTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATTCTTG 1429
          |||||
Db      241  CATTTCATCATTGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATTCTTG 300
          |||||
Qy      1430 TAATATTAAAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTAT 1489
          |||||
Db      301  TAATATTAAAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTAT 360
          |||||
Qy      1490 TTGAAAATTATATTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543
          |||||
Db      361  TTGAAAATTATATTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 414
          |||||

```

# RESULT 8

BB847918

EOCUS BB847918 422 bp mRNA linear EST 26-NOV-2001

DEFINITION BB847918 RIKEN full-length enriched, adult male kidney Mus musculus cDNA clone F530201F11 5', mRNA sequence.

ACCESSION BB847918

VERSION BB847918.1 GI:17086293

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 422)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and

181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 240

Db 196 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT 255  
 Qy 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300  
 Db 256 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 315  
 Qy 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360  
 Db 316 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 375  
 Qy 361 CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGT 407  
 Db 376 CCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGGTCATGAAGT 422

RESULT 9

BB864882

LOCUS BB864882 420 bp mRNA linear EST 27-NOV-2001

DEFINITION BB864882 RIKEN full-length enriched, RCB-1283 B16 melanoma cDNA Mus musculus cDNA clone G430047C11 5', mRNA sequence.

ACCESSION BB864882

VERSION BB864882.1 GI:17111092

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 420)

AUTHORS Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and



Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES                      Location/Qualifiers  
    source                      1. .420  
                                /organism="Mus musculus"  
                                /mol\_type="mRNA"  
                                /db\_xref="taxon:10090"  
                                /clone="G430047C11"  
                                /cell\_line="RCB-1283 B16 melanoma"  
                                /clone\_lib="RIKEN full-length enriched, RCB-1283 B16  
                                melanoma cDNA"

BASE COUNT                      102 a                      103 c                      87 g                      128 t  
ORIGIN

Query Match                      25.2%;    Score 388.4;    DB 10;    Length 420;  
Best Local Similarity    99.5%;    Pred. No. 7.3e-70;  
Matches 400;    Conservative    0;    Mismatches    1;    Indels    1;    Gaps    1;

```
Qy      1 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTATC 60
      |||
Db      19 GCTCCTGGCAGAGTTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTATC 78

Qy      61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120
      |||
Db      79 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 138

Qy     121 TGCAATCGAGTTCATTTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCTT 180
      |||
Db     139 TGCAATCGAGTTCATTTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCTT 198

Qy     181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT 240
      |||
Db     199 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT 258

Qy     241 TGCTTTCTGTGCACCCTTCCCATCTTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300
      |||
Db     259 TGCTTTCTGTGCACCCTTCCCATCTTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 318

Qy     301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360
      |||
Db     319 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 378

Qy     361 CCTCTTCCTCACTTTTCATTAG-CATGGACCGATATCTGCTCA 401
      |||
Db     379 CCTCTTCCTCACTTTTCATTAGCCATGGACCGATATCTGCTCA 420
```

RESULT 10

BB778587

LOCUS BB778587 426 bp mRNA linear EST 15-NOV-2001

DEFINITION BB778587 RIKEN full-length enriched, RCB-1283 B16 melanoma cDNA Mus musculus cDNA clone G430047C11 3', mRNA sequence.

ACCESSION BB778587

VERSION BB778587.1 GI:16939287

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 426)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES Location/Qualifiers

source 1..426

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"  
 /clone="G430047C11"  
 /cell\_line="RCB-1283 B16 melanoma"  
 /clone\_lib="RIKEN full-length enriched, RCB-1283 B16  
 melanoma cDNA"

BASE COUNT 153 a 58 c 76 g 139 t  
 ORIGIN

Query Match 24.9%; Score 384.8; DB 10; Length 426;  
 Best Local Similarity 98.8%; Pred. No. 4.1e-69;  
 Matches 419; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

```

Qy      1123 GGGCTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAG 1182
          |||
Db      3 GGGCTGGAGTACAAGCTGGGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAG 62

Qy      1183 GTTATACCCAGAGTATGGAAAAAATAA-GGCATGAGAAAGCATTGACATCTTCACCTAAG 1241
          |||
Db      63 GTTATACCCAGAGTATGGAAAAAATAAGGGCATGAAAAGCATTGACATCTTCACCTAAG 122

Qy      1242 AACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGG 1301
          |||
Db      123 AACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTTG 182

Qy      1302 AAATTTTAAGACCTCTTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAAT 1361
          |||
Db      183 AAATTTTAAGACCTCTTTTTCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAAT 242

Qy      1362 GCTGAATGCATTTCATCATTGGTCA-3GTCGATAAGCGTGTTTCTGAAATAGTCTTATTT 1420
          |||
Db      243 GCTGAATGCATTTCATCATTGGTCAAGCGTGTTTCTGAAATAGTCTTATTT 302

Qy      1421 TTATTCTTGTAATATTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAG 1480
          |||
Db      303 TTATTCTTGTAATATTAAATTTATGTGAAAAATGAATATAATTCAATGTACAACATTAG 362

Qy      1481 ATTTTCTA-TTTGAAAATTATATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAA 1539
          |||
Db      363 ATTTTCTAGTTTGAAAATTATATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAA 422

Qy      1540 TATA 1543
          |||
Db      423 TATA 426
  
```

RESULT 11

BB739482

LOCUS BB739482 396 bp mRNA linear EST 15-OCT-2001

DEFINITION BB739482 RIKEN full-length enriched, 6 days neonate spleen Mus  
 musculus cDNA clone F430113M16 3', mRNA sequence.

ACCESSION BB739482

VERSION BB739482.1 GI:16142487

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 396)  
 AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

FEATURES Location/Qualifiers  
 source 1..396  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="F430113M16"  
 /tissue\_type="spleen"  
 /dev\_stage="6 days neonate"  
 /clone\_lib="RIKEN full-length enriched, 6 days neonate spleen"

BASE COUNT 142 a 52 c 62 g 140 t  
 ORIGIN

Query Match 24.7%; Score 380.4; DB 10; Length 396;  
 Best Local Similarity 99.7%; Pred. No. 3.3e-68;

Matches 381; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1162 AGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAG 1221
          |||
Db      1   AGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATAAGGCATGAGAAAG 60

Qy      1222 CATTGACATCTTCACCTTAAGAACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACA 1281
          |||
Db      61   CATTGACATCTTCACCTTAAGAACTGAACAAAAGAGAACAAATATTGTCAATGTTTGGACA 120

Qy      1282 CTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTTCTATCAGTGTAAGGAATA 1341
          |||
Db      121  CTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTTCTATCAGTGTAAGGAATA 180

Qy      1342 CAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGT 1401
          |||
Db      181  CAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGGTCGATAAGCGTGT 240

Qy      1402 TTCTGAAATAGTCTTATTTTATTCTTGTAATATTTAAATTTATGTGAAAAATGAATATA 1461
          |||
Db      241  TTCTGAAATAGTCTTATTTTATTCTTGTAATATTTAAATTTATGTGAAAAATGAATATA 300

Qy      1462 ATTCAATGTACAACATTAGATTTTCTATTTGAAAATTATATTTCTTGAAAAATAACTGC 1521
          |||
Db      301  ATTCAATTTACAACATTAGATTTTCTATTTGAAAATTATATTTCTTGAAAAATAACTGC 360

Qy      1522 TGTGCCTAAATAAATCAATATA 1543
          |||
Db      361  TGTGCCTAAATAAATCAATATA 382
```

#### RESULT 12

AI649254/c

LOCUS AI649254 367 bp mRNA linear EST 30-APR-1999

DEFINITION uk27c10.x1 Sugano mouse kidney mkia Mus musculus cDNA clone

IMAGE:1970226 3', mRNA sequence.

ACCESSION AI649254

VERSION AI649254.1 GI:4730088

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 367)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished

COMMENT Other ESTs: uk27c10.y1

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:986966

This clone was previously sequenced on the 5' end only, this new  
data is from the 3' end

Seq primer: custom primer used

High quality sequence stop: 353.

#### FEATURES

source

Location/Qualifiers

1. .367

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL"

/db\_xref="taxon:10090"

/clone="IMAGE:1970226"

/sex="female"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="Sugano mouse kidney mkia"

/note="Organ: kidney; Vector: pME18S-FL3; Site\_1: DraIII  
(CACTGTGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCCTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor [TGTGGCCTACTGG], digested  
and cloned into distinct DraIII sites of the pME18S-FL3  
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end  
primer CGACCTGCAGCTCGAGCACA."

BASE COUNT

106 a 73 c 71 g 117 t

ORIGIN

Query Match 23.6%; Score 363.8; DB 9; Length 367;

Best Local Similarity 99.5%; Pred. No. 8.9e-65;

Matches 365; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1035 CACTTGATAAACAGTGCTGTGCAGTTGAGTTTTAACTAAGTAAACCACCATTCTAGGCT 1094
      |||
Db 367 CACTTGATAAACAGTGCTGTGCAGTTGAGTTTTAACTAAGTAAACCACCATTCTACGCT 308

Qy 1095 TTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTGGGTCCACATGAATC 1154
      |||
Db 307 TTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTGGGTCCACATGAATC 248

Qy 1155 AGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATAAGGCAT 1214
      |||
Db 247 AGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATAAGGCAT 188

Qy 1215 GAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACAAATATTGTCAATGT 1274
      |||
Db 187 GAGAAAGCATTGACATCTTCACTTAAGATCTGAACAAAAGAGAACAAATATTGTCAATGT 128

Qy 1275 TTGGACACTTAGGATCTGAAATCTTGGAATTTTAAAGACCTCTTTTCTATCAGTGTAATA 1334
      |||
```

Db 127 TTGGACACTTAGGATCTGAAATCTTGCAAATTTTAAGACCTCTTTTCTATCAGTGATAA 68

Qy 1335 AGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGGTCGATA 1394  
 |||

Db 67 AGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGGTCGATA 8

Qy 1395 AGCGTGT 1401  
 |||

Db 7 AGCGTGT 1

# RESULT 13

BB645274

LOCUS BB645274 636 bp mRNA linear EST 31-AUG-2001

DEFINITION BB645274 RIKEN full-length enriched, 4 days neonate male adipose  
 Mus musculus cDNA clone B430012021 5', mRNA sequence.

ACCESSION BB645274

VERSION BB645274.1 GI:15402306

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 636)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
 Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
 Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
 Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,  
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
 Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
 Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.

Func. Genomics 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

#### FEATURES

source

Location/Qualifiers

1. .636

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="B430012021"

/sex="male"

/tissue\_type="adipose"

/dev\_stage="4 days neonate"

/lab\_host="DH10B"

/clone\_lib="RIKEN full-length enriched, 4 days neonate male adipose"

/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'

GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 229.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 170 a 118 c 140 g 207 t 1 others  
ORIGIN

Query Match 23.2%; Score 357.6; DB 10; Length 636;  
Best Local Similarity 91.7%; Pred. No. 1.8e-63;  
Matches 389; Conservative 0; Mismatches 34; Indels 1; Gaps 1;

```
Qy      1 GCTCCTGGCAGAGTTTTCTGTCTGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60
      |||
Db      20 GCTCCTGGCAGAGTTTTCTGTCTGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 79

Qy     61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120
      |||
Db     80 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGGATTTTA 139

Qy    121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTT 180
      |||
Db    140 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGGGGTGTTCGGGTACCTGTT 199
```



Qy 181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 240  
 |||  
 Db 200 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACTT 259  
 Qy 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300  
 |||  
 Db 260 TGCTTTCCTGGGCACCCTTCCCATCCTGATAAAGAGTTTGGCCAATGATAAGGGGACCTA 319  
 Qy 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACC-AGCA 359  
 |||  
 Db 320 TGGAGATGTTCTTTGGATAAGCAACCGATATGGGCTTAACACCAACCTTTAATCCAAGCT 379  
 Qy 360 TCCTCTTCCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAG 419  
 |||  
 Db 380 TACTTTTACTTACTTTATTTAGCATGGACCGATATTTGTTTATGAAAGTGCCTTTTTCG 439  
 Qy 420 AACA 423  
 |||  
 Db 440 AAAA 443

# RESULT 14

BB846608

LOCUS BB846608 416 bp mRNA linear EST 26-NOV-2001

DEFINITION BB846608 RIKEN full-length enriched, adult male kidney *Mus musculus* cDNA clone F530003I24 5', mRNA sequence.

ACCESSION BB846608

VERSION BB846608.1 GI:17084983

KEYWORDS EST.

SOURCE *Mus musculus* (house mouse)

ORGANISM *Mus musculus*

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; *Mus*.

REFERENCE 1 (bases 1 to 416)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished

COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

#### FEATURES

source

Location/Qualifiers

1. .416

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="F530003I24"

/sex="male"

/tissue\_type="kidney"

/dev\_stage="adult"

/lab\_host="SOLR"

/clone\_lib="RIKEN full-length enriched, adult male kidney"

/note="Site\_1: XhoI; Site\_2: SstI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer [5'

GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the

primer adapter of sequence [5'

GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTAATTAACCCCCCCCCC 3'].

cDNA was cleaved with XhoI and SstI. "

BASE COUNT 107 a 93 c 87 g 129 t

ORIGIN

Query Match 23.0%; Score 354.2; DB 10; Length 416;

Best Local Similarity 97.3%; Pred. No. 8.7e-63;

Matches 392; Conservative 0; Mismatches 8; Indels 3; Gaps 3;

QY 1 GCTCCTGGCAGAGTTTTCTGTGTCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60

|||||

Db 16 GCTCCTGGCAGAGTTTTCTGTGTCGAGACAGAAGCCGAAAGCAGAATGGCACAGAATTTATC 75

QY 61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120

|||||

Db 76 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 135

Qy 121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTT 180  
 |||||  
 Db 136 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCTT 195  
 Qy 181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACTTTCCATCTCTGACTT 240  
 |||||  
 Db 196 CTGCATGAAGAACTGGAAAAGCAGCAATGTCTATCTTTTAACTTT-CATCTCTGACTT 254  
 Qy 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300  
 |||||  
 Db 255 TGCTTTCCTGTGCACCCTT-CCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 313  
 Qy 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360  
 |||||  
 Db 314 TGGAGATGTTCTATGTATAAGCAACCGATATGTGGTTTACAACAACCTCTAAACCAGCAT 373  
 Qy 361 CCTCTTCCTCACTTTCATTAG-CATGGACCGATATCTGCTCAT 402  
 |||||  
 Db 374 CCTCTTCCTCACTTTCATTAGCCATGGACCGATATCTGCTCAT 416

# RESULT 15

BY368584

LOCUS BY368584 408 bp mRNA linear EST 12-DEC-2002

DEFINITION BY368584 RIKEN full-length enriched, 6 days neonate spleen Mus musculus cDNA clone F430110C01 3', mRNA sequence.

ACCESSION BY368584

VERSION BY368584.1 GI:26598072

KEYWORDS EST

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 408)

AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perteau,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,P.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.

, Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**JOURNAL** Nature 420, 563-573 (2002)

**MEDLINE** 22354683

**PUBMED** 12466851

**COMMENT** Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES** Location/Qualifiers

source 1. .408  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="F430110C01"  
 /tissue\_type="spleen"  
 /dev\_stage="6 days neonate"  
 /clone\_lib="RIKEN full-length enriched, 6 days neonate spleen"

**BASE COUNT** 145 a 58 c 63 g 141 t 1 others

**ORIGIN**

Query Match 22.7%; Score 350.6; DB 13; Length 408;

Best Local Similarity 98.0%; Pred. No. 4.8e-62;  
Matches 386; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```
Qy      1153 TCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGGAAAAAATAAGGC 1212
          |||||
Db      1 TCAGAAGCCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGNAAAAAATAAGGC 60

Qy      1213 ATGAGAAAGCA-TTGACATCTTCACTTAAGAACTGAACAAAAGAGAACAAATA-TTGTCA 1270
          |||||
Db      61 ATGAGAAAGCAGTTGACATCTTCACTTAAGAACTGAACAAAAGAGAACAAATAGTTCTCA 120

Qy      1271 ATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTTCTATCAGTG 1330
          |||||
Db      121 ATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTTCTATCAGTG 180

Qy      1331 TAAAAGGAATACAAGATAG-CTAGTTGCAAATGCTGAATGCATTTTCATCATTGGTCAGGT 1389
          |||||
Db      181 TAAAAGGAATACAAGATAGCCTAGTTCCAAATGCTGAATGCATTTTCATCATTGGTCAGGT 240

Qy      1390 CGATAAGCGTGTCTGAAATAGTCTTATTTTTATTCTTGTAATATTAAAATTTATGTGA 1449
          |||||
Db      241 CGATAAGCGTGTCTGAAATAGTCTTATTTTTATTCTTGTAATATTAAAATTTATGTGA 300

Qy      1450 AAAATGAATAAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTATATTTCTTGA 1509
          |||||
Db      301 AAAATGAATATAATTCAATTTACAACATTAGATTTTCTATTTGAAAATTATATTTCTTGA 360

Qy      1516 AAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543
          |||||
Db      361 AAAAATAACTGCTGTGCCTAAATAAATCAATATA 394
```

Search completed: December 14, 2003, 17:26:57  
Job time : 3525 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 14, 2003, 10:09:19 ; Search time 450 Seconds  
(without alignments)  
9256.073 Million cell updates/sec

Title: US-09-891-138A-1  
Perfect score: 1543  
Sequence: 1 gctcctggcagagttttctg.....tgcctaaataaatcaatata 1543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:\*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

8						
Result	Query					
No.	Score	Match	Length	DB	ID	Description
1	1543	100.0	1543	24	ABK12957	DNA sequence of mo
2	592.4	38.4	1005	21	AAD01135	Human orphan G pro
3	592.4	38.4	1005	21	AAA46036	Human G protein co
4	592.4	38.4	1380	25	ABZ42542	Human purinergic r
5	592.4	38.4	1436	24	ABL90790	Human polynucleoti
6	592.4	38.4	1473	25	ACC46165	Human dithp recept
7	592.4	38.4	1542	24	AAD24958	Human G-protein co
8	590.8	38.3	1338	25	ABS57291	cDNA encoding huma
9	589.2	38.2	1996	18	AAT71900	Human purinergic r
10	587.2	38.1	1428	18	AAT75146	Human ATP receptor
11	580.4	37.6	1385	21	AAC81122	Human secreted pro
12	126.6	8.2	1011	24	AAL43942	Human G protein-co
13	126.6	8.2	1014	22	AAS07948	Human cDNA encodin
14	126.6	8.2	1014	24	ABZ42876	Human GPCR polynuc
15	126.6	8.2	1014	24	ABN85630	Human P2Y-like rec
16	126.6	8.2	1014	24	ABK11381	Human DNA encoding
17	126.6	8.2	1014	24	AAK98323	Human purinergic-r
18	126.6	8.2	1014	25	ABZ59170	Human TGR164 prote
19	126.6	8.2	1014	25	ABZ42582	Human G protein-co
20	126.6	8.2	1029	25	AAD50885	TARZAN DNA compris
21	126.6	8.2	1081	24	ABS59232	Human G-protein co
22	126.6	8.2	1179	25	AAD50883	Human TARZAN cDNA
23	126.6	8.2	1288	24	ABL56197	Human P2Y1-li enco
24	126.6	8.2	1729	22	AAS08362	Human cDNA encodin
25	126.6	8.2	1729	23	ABV24026	Human prostate exp
26	126.6	8.2	1729	23	ABV25767	Human prostate exp
27	126.6	8.2	1729	23	ABV29909	Human prostate exp
28	126.6	8.2	1729	23	ABV30024	Human prostate exp
29	126.6	8.2	1797	25	AAD50882	Human TARZAN cDNA.
30	126.6	8.2	5435	24	ABL56198	Human P2Y1-li enco
31	126.6	8.2	9905	24	AAK98324	Human purinergic-r
32	125	8.1	1014	24	ABQ78847	Human G-protein co
33	125	8.1	1014	24	AAD34278	Human AXOR89 (G-pr
34	125	8.1	1014	24	AAD26370	Human G-protein co
35	125	8.1	1560	24	ABS51730	Human novel polynu
36	125	8.1	1851	24	ABS51678	DNA encoding human
37	119.2	7.7	1020	22	AAH51011	Human nGPCR54 codi
38	119.2	7.7	1020	24	ABS70244	DNA encoding human
39	119	7.7	1313	22	AAK52430	Human polynucleoti
c 40	112.6	7.3	740	23	ABV15662	Human prostate exp
41	109	7.1	1014	25	AAD50884	Mouse TARZAN cDNA.
42	109	7.1	7399	25	AAD50886	Mouse TARZAN genom
43	104	6.7	1020	24	ABQ79300	Human GPCR designa
44	104	6.7	1076	24	AAD29667	Human G-protein co
45	103.2	6.7	6721	24	AAS18600	Purinergic recepto

#### ALIGNMENTS

RESULT 1

ABK12957

ID ABK12957 standard; DNA; 1543 BP.

XX

AC ABK12957;

XX

DT 09-APR-2002 (first entry)

XX

DE DNA sequence of mouse G-protein coupled receptor TGR18 gene.

XX

KW Mouse; G-protein coupled; receptor; GPCR; TGR18; kidney disease;  
KW signal transduction modulator; cerebral cavernous malformation;  
KW hyperlipidemia; obesity; dyslexia; cardiac myxoma; renal failure;  
KW nephritis; hypertension; liver disease; cirrhosis; blood disorder;  
KW spleen-associated disorder; immune disorder; gene; ds.

XX

OS Mus sp.

XX

FH Key Location/Qualifiers

FT CDS 44..997

FT /\*tag= a

FT /product= "Mouse G-protein coupled receptor TGR18"

XX

PN WO200200719-A2.

XX

PD 03-JAN-2002.

XX

PF 25-JUN-2001; 2001WO-US20363.

XX

PR 23-JUN-2000; 2000US-213461P.

XX

PA (TULA-) TULARIK INC.

XX

PI Lin DC, Zhao J, Chen J, Cutler G;

XX

DR WPI; 2002-147880/19.

DR P-PSDB; AAU74904.

XX

PT New G-protein coupled receptor polypeptides, useful for identifying  
PT modulators of signal transduction for treating kidney disease,  
PT hyperlipidemia, obesity, dyslexia and cardiac myxoma -

XX

PS Claim 18; Page 58; 78pp; English.

XX

CC The present invention relates to a new G-protein coupled receptor (GPCR)  
CC polypeptide comprising greater than 70% amino acid sequence identity to  
CC the amino acid sequence of human GPCRs TGR62, TGR21, TGR130.1, TGR130.2,  
CC human TGR213 or TGR92, 80% amino acid sequence identity to mouse TGR18  
CC or 90% amino acid sequence identity to human novel edg receptor protein,  
CC as defined in the specification. The GPCR covalently linked to a solid  
CC phase is useful for identifying a compound that modulates signal  
CC transduction. The identified compounds are useful for treating  
CC kidney disease, cerebral cavernous malformations, hyperlipidemia,  
CC obesity, dyslexia and cardiac myxoma. The molecules of the invention are  
CC useful for diagnosing disorders or conditions such as kidney-related  
CC conditions or diseases such as renal failure, nephritis, nephrotic



CC syndrome, asymptomatic urinary abnormalities, renal tubule defects,  
CC hypertension and nephrolithiasis, liver-related disease or condition  
CC e.g. cirrhosis, infiltrations, lesions, functional disorders and jaundice  
CC and spleen-associated disorders or conditions e.g. splenic enlargement,  
CC immune disorders, blood disorders and others. Modulation of the  
CC polypeptide of the invention is useful to treat or prevent any of the  
CC above conditions or diseases. The present nucleic acid sequence encodes  
CC the mouse GPCR TGR18 protein of the invention. This sequence encodes one  
CC of seven novel G protein coupled receptors of the invention (ABK12957-  
CC ABK12964).

XX

SQ Sequence 1543 BP; 438 A; 352 C; 293 G; 460 T; 0 other;

Query Match 100.0%; Score 1543; DB 24; Length 1543;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCTCCTGGCAGAGTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60  
|  
Db 1 GCTCCTGGCAGAGTTTCTGTGCGAGACAGAAGCCGACAGCAGAATGGCACAGAATTTATC 60

Qy 61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120  
|  
Db 61 TTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTTA 120

Qy 121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCCGGCTACCTCTT 180  
|  
Db 121 TGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCCGGCTACCTCTT 180

Qy 181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACCTTTCCATCTCTGACTT 240  
|  
Db 181 CTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAACCTTTCCATCTCTGACTT 240

Qy 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300  
|  
Db 241 TGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAATGATAAGGGGACCTA 300

Qy 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360  
|  
Db 301 TGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCAGCAT 360

Qy 361 CCTCTTCCTCACTTTTATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA 420  
|  
Db 361 CCTCTTCCTCACTTTTATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCCGAGA 420

Qy 421 ACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT 480  
|  
Db 421 ACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCTTAGT 480

Qy 481 GACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAG 540  
|  
Db 481 GACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGGGCAG 540

Qy 541 TAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTG 600  
|  
Db 541 TAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCCTCTG 600

Qy	601	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Db	601	CCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACAAGAT	660
Qy	661	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Db	661	GGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACAAACC	720
Qy	721	CCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Db	721	CCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATCATAT	780
Qy	781	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Db	781	CATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTTGGCCACAAGGATGTACACAGAA	840
Qy	841	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAA	900
Db	841	GGCCATCAAATCTATATACACACTGACACGGCCTCTGGCCTTTCTGAACAGTGCCATCAA	900
Qy	901	TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAG	960
Db	901	TCCCATCTTCTACTTCCTCATGGGAGACCATTACAGAGAGATGCTGATTAGTAAGTTCAG	960
Qy	961	ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT	1020
Db	961	ACAATACTTCAAGTCCCTTACATCCTTCAGGACATGAGCTGCTGGATGCAGGTCTTCACT	1020
Qy	1021	CAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTTAACTAAGTAAACC	1080
Db	1021	CAGCCAAAATGAGACACTTGATAAACAGTGCTGTGCAGTTGAGTTTTAACTAAGTAAACC	1080
Qy	1081	ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTG	1140
Db	1081	ACCATTTCTAGGCTTTAGCTTTCCACCATCCTCCAACCCCCAGGGCTGGAGTACAAGCTG	1140
Qy	1141	GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG	1200
Db	1141	GGTCCACATGAATCAGAAGGCAGCTCTCTGTTCTGATTTTAGGTTATACCCAGAGTATGG	1200
Qy	1201	AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA	1260
Db	1201	AAAAAATAAGGCATGAGAAAGCATTGACATCTTCACTTAAGAACTGAACAAAAGAGAACA	1260
Qy	1261	AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTT	1320
Db	1261	AATATTGTCAATGTTTGGACACTTAGGATCTGAAATCTTGAAATTTTAAGACCTCTTTT	1320
Qy	1321	TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT	1380
Db	1321	TCTATCAGTGTAAGGAATACAAGATAGCTAGTTGCAAATGCTGAATGCATTTTCATCAT	1380
Qy	1381	TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATTCTTGTAATATTAAAA	1440
Db	1381	TGGTCAGGTCGATAAGCGTGTTTCTGAAATAGTCTTATTTTTATTCTTGTAATATTAAAA	1440
Qy	1441	TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTAT	1500

```

      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1441 TTTATGTGAAAAATGAATATAATTCAATGTACAACATTAGATTTTCTATTTGAAAATTAT 1500

Qy      1501 ATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1501 ATTTCTTGAAAAAATAACTGCTGTGCCTAAATAAATCAATATA 1543

```

RESULT 2

AAD01135

ID AAD01135 standard; cDNA; 1005 BP.

XX

AC AAD01135;

XX

DT 02-NOV-2000 (first entry)

XX

DE Human orphan G protein-coupled receptor hCHN10 cDNA.

XX

KW Human; orphan G protein-coupled receptor; GPCR; hCHN10; drug screening;

KW transmembrane receptor; expressed sequence tag; EST; signal cascade; ss.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	CDS	1..1005
----	-----	---------

FT		/*tag= a
----	--	----------

FT		/product= "hCHN10"
----	--	--------------------

FT		/note= "Human orphan G protein-coupled receptor"
----	--	--

XX

PN WO200031258-A2.

XX

PD 02-JUN-2000.

XX

PF 13-OCT-1999; 99WO-US23687.

XX

PR 20-NOV-1998; 98US-0109213.

PR 16-FEB-1999; 99US-0120416.

PR 26-FEB-1999; 99US-0121852.

PR 12-MAR-1999; 99US-0123946.

PR 12-MAR-1999; 99US-0123949.

PR 28-MAY-1999; 99US-0136436.

PR 28-MAY-1999; 99US-0136437.

PR 28-MAY-1999; 99US-0136439.

PR 28-MAY-1999; 99US-0136567.

PR 28-MAY-1999; 99US-0137127.

PR 28-MAY-1999; 99US-0137131.

PR 29-JUN-1999; 99US-0141448.

PR 29-SEP-1999; 99US-0156555.

PR 29-SEP-1999; 99US-0156633.

PR 29-SEP-1999; 99US-0156634.

PR 29-SEP-1999; 99US-0156653.

PR 01-OCT-1999; 99US-0157280.

PR 01-OCT-1999; 99US-0157281.

PR 01-OCT-1999; 99US-0157282.

PR 01-OCT-1999; 99US-0157293.

PR 01-OCT-1999; 99US-0157294.

PR 12-OCT-1999; 99US-0416760.

XX

XX

XX

DR

XX

XX

XX

XX

Db

Db

Db

Db

Dh

Ov



KW Human; G protein coupled receptor; GPCR; transmembrane receptor;  
 KW identification; agonist; screening; therapeutic; pharmaceutical;  
 KW mutant; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200022131-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 13-OCT-1999; 99WO-US24065.  
 XX  
 PR 13-OCT-1998; 98US-0170496.  
 PR 12-NOV-1998; 98US-0108029.  
 PR 20-NOV-1998; 98US-0109213.  
 PR 27-NOV-1998; 98US-0110060.  
 PR 16-FEB-1999; 99US-0120416.  
 PR 26-FEB-1999; 99US-0121852.  
 PR 12-MAR-1999; 99US-0123944.  
 PR 12-MAR-1999; 99US-0123945.  
 PR 12-MAR-1999; 99US-0123946.  
 PR 12-MAR-1999; 99US-0123948.  
 PR 12-MAR-1999; 99US-0123949.  
 PR 12-MAR-1999; 99US-0123951.  
 PR 28-MAY-1999; 99US-0136436.  
 PR 28-MAY-1999; 99US-0136437.  
 PR 28-MAY-1999; 99US-0136439.  
 PR 28-MAY-1999; 99US-0137127.  
 PR 28-MAY-1999; 99US-0137131.  
 PR 28-MAY-1999; 99US-0137567.  
 PR 30-JUN-1999; 99US-0141448.  
 PR 27-AUG-1999; 99US-0151114.  
 PR 03-SEP-1999; 99US-0152524.  
 PR 29-SEP-1999; 99US-0156633.  
 PR 29-SEP-1999; 99US-0156555.  
 PR 29-SEP-1999; 99US-0156634.  
 XX  
 PA (AREN-) ARENA PHARM INC.  
 XX  
 PI Behan DP, Lehmann-Bruinsma K, Chalmers DT, Chen R, Dang HT;  
 PI Gore M, Liaw CW, Lin I, Lowitz K, White C;  
 XX  
 DR WPI; 2000-317986/27.  
 DR P-PSDB; AAB02842.  
 XX  
 PT Non-endogenous, human G protein-coupled receptors for screening  
 PT receptor, inverse or partial agonists useful as therapeutic agents -  
 XX  
 PS Example 1; Page 116; 187pp; English.  
 XX  
 CC The present invention describes transmembrane receptors, preferably  
 CC human G protein coupled receptors (GPCR), for which the endogenous  
 CC ligand is unknown (orphan GPCR receptors). More specifically the present  
 CC invention relates to non-endogenous, constitutively activated versions  
 CC of a human GPCR. These non-endogenous human GPCRs can be useful for  
 CC the direct identification of candidate compounds as receptors agonists,  
 CC inverse agonists or partial agonists for use as pharmaceutical agents.

CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in  
CC the exemplification of the present invention.

XX

SQ Sequence 1005 BP; 248 A; 236 C; 196 G; 325 T; 0 other;

Query Match 38.4%; Score 592.4; DB 21; Length 1005;  
Best Local Similarity 75.5%; Pred. No. 3.2e-140;  
Matches 750; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      8 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 67

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     68 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 127

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     128 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 187

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     188 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCATGCTGATAAGGAGTT 247

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     248 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 307

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     308 ATGCCAACCTCTATACCAGCATTCTCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 367

Qy     399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     368 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 427

Qy     459 CGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     428 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 487

Qy     519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     488 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT 547

Qy     579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     548 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTTGTGA 607

Qy     639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGCCAGCAGCAAGCAA 698
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     608 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 667

Qy     699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     668 CTGCTCTGCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 727
```





DR WPI; 2003-046718/04.

DR P-PSDB; ABP81696.

XX

PT New isolated antigenic peptides e.g., for G protein-coupled receptors  
PT (GPCR), useful for diagnosing and designing drugs for treating  
PT conditions in which GPCRs are involved, e.g. AIDS, Alzheimer's disease,  
PT cancer or autoimmune diseases -

XX

PS Disclosure; Fig 1; 523pp; English.

XX

CC The present invention describes antigenic peptides (I) comprising:  
CC (a) any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino  
CC acids. Also described: (1) an assay for the detection of a particular  
CC G protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;  
CC and (2) an isolated antibody having high specificity and high affinity  
CC or avidity for a particular GPCR. (I) can be used as GPCR modulators and  
CC in gene therapy. The antigenic peptides for GPCRs are useful in detecting  
CC an antibody against a particular GPCR, and in the production of specific  
CC antibodies. The peptides and antibodies are also useful for detecting the  
CC presence or absence of corresponding GPCRs. The antigenic peptides for  
CC GPCRs and antibodies are useful for diagnosing and designing drugs for  
CC treating immune-related diseases, growth-related diseases, cell  
CC regeneration-related disease, immunological-related cell proliferative  
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,  
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,  
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute  
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host  
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,  
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory  
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,  
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or  
CC any other disorder in which GPCRs are involved. The antibodies may be  
CC used in immunoassays and immunodiagnosis. ABZ42523 to ABZ42869 encode  
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the  
CC exemplification of the present invention.

XX

SQ Sequence 1380 BP; 383 A; 294 C; 274 G; 429 T; 0 other;

Query Match 38.4%; Score 592.4; DB 25; Length 1380;  
Best Local Similarity 75.3%; Pred. No. 3.7e-140;  
Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

Qy	39	GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA	98
Db	50	GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA	109
Qy	99	AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA	158
Db	110	AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAATACCA	169
Qy	159	CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT	218
Db	170	TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT	229
Qy	219	TTAACCTTTCCATCTCTGACTTTGCTTTTCTGTGCACCCCTTCCCATCTGATAAAGAGTT	278
Db	230	TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCCTCCCCTGCTGATAAGGAGTT	289



XX  
 AC ABL90790;  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 1352.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US16450.  
 XX  
 PR 19-MAY-2000; 2000US-205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-122018/16.  
 DR P-PSDB; ABB90381.  
 XX  
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 4; SEQ ID NO 1352; 2081pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1436 BP; 397 A; 309 C; 289 G; 441 T; 0 other;

Query Match 38.4%; Score 592.4; DB 24; Length 1436;  
Best Local Similarity 75.3%; Pred. No. 3.7e-140;  
Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      100 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 159

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      160 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAATAACCA 219

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     220 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 279

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     280 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 339

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     340 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 399

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     400 ATGCCAACCTCTATACCAGCATCTCTTTCTCACTTTATCAGCATAGATCGATACTTGA 459

Qy     399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     460 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 519

Qy     459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     520 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 579

Qy     519 CTGTCCCAAAAGAAGAGGGCAGTAACCTGCATCGACTATGCAAGTTCGGAACCCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     580 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCGAGAGCCCCAACT 639

Qy     579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     640 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA 699

Qy     639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGCCAGCAGCAAGCAA 698
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     700 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 759

Qy     699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTA 758
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     760 CTGCTCTGCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGAATCTTCTCTG 819

Qy     759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     820 TGCTTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 879
```



PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX

DR WPI; 2003-129518/12.

DR P-PSDB; ABR41222.

XX

PT Novel human diagnostic and therapeutic polypeptide useful for  
PT identifying test compound which specifically binds to a polypeptide  
PT encoded by human diagnostic and therapeutic polynucleotide, and to  
PT induce antibodies -

XX

PS Claim 2; SEQ ID No 86; 591pp; English.

XX

CC The invention relates to novel human diagnostic and therapeutic  
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their  
CC encoded proteins (DITHP; ABR41136-ABR41812). The invention also relates  
CC to polynucleotide sequences at least 90% identical to the dithp cDNA  
CC sequences of the invention; recombinant vectors, host cells and  
CC transgenic organisms comprising a dithp nucleic acid sequence; the  
CC recombinant production of DITHP proteins; antibodies specific for DITHP  
CC proteins; microarrays comprising dithp nucleic acid sequences; methods  
CC of detecting dithp nucleotide and protein sequences; methods of screening  
CC for compounds which specifically bind a DITHP protein; and methods of  
CC assessing the toxicity of test compounds using a dithp hybridisation  
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the  
CC diagnosis of a wide variety of conditions including cancer and other cell  
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,  
CC viral, fungal or parasitic infections; hormonal disorders; metabolic  
CC disorders; neurological disorders; gastrointestinal disorders; transport  
CC disorders; and connective tissue disorders. They may also be used to  
CC screen for modulators of protein activity or gene expression. DITHP  
CC proteins can additionally be used in analysis of the proteome of a tissue  
CC or cell type and to induce antibodies. The dithp nucleic acids are  
CC additionally useful in somatic or germline gene therapy of the disorders  
CC mentioned above, as a source of antisense sequences, as a source of  
CC probes and primers, in genotyping and identification of individuals, in  
CC the generation of transgenic animal models of human disease or knock in  
CC humanised animals, in toxicological testing, and in transcript imaging.  
CC The present sequence represents a dithp cDNA encoding a DITHP protein  
CC which has receptor activity.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1473 BP; 403 A; 320 C; 303 G; 447 T; 0 other;

Query Match 38.4%; Score 592.4; DB 25; Length 1473;  
Best Local Similarity 75.3%; Pred. No. 3.8e-140;  
Matches 764; Conservative 0; Mismatches 246; Indels 4; Gaps 2;

Qy 39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98  
| | ||||| ||| | ||| | || |||| ||||| | || |  
Db 119 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 178  
  
Qy 99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA 158

Db 179 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGAAATACCA 238  
 Qy 159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218  
 Db 239 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 298  
 Qy 219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278  
 Db 299 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCATGCTGATAAGGAGTT 358  
 Qy 279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338  
 Db 359 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 418  
 Qy 339 ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC 398  
 Db 419 ATGCCAACCTCTATACCAGCATCTCTTTCTCACTTTTATCAGCATAGATCGATACTTGA 478  
 Qy 399 TCATGAAGTACCCTTTCCGAGAACACCTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458  
 Db 479 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 538  
 Qy 459 CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518  
 Db 539 CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 598  
 Qy 519 CTGTCCCAAGAAGAGGGCAGTAACCTGCATCGACTATGCAAGTTCTGGAAACCCTGAAC 578  
 Db 599 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGAAGTTCTGGAGACCCCAACT 658  
 Qy 579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638  
 Db 659 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTTGTGA 718  
 Qy 639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698  
 Db 719 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 778  
 Qy 699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758  
 Db 779 CTGCTCTGCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 838  
 Qy 759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818  
 Db 839 TGCTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 898  
 Qy 819 G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC 875  
 Db 899 GGAAGCAGTATCAGTGCATCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTT 958  
 Qy 876 TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCTCATGGGAGACCATTACA 935  
 Db 959 TGGCCTTTCTGAACAGTGTCAACCCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA 1018  
 Qy 936 GAGAGATGCTGATTAGTAAGTTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT 995

Db 1019 GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT 1078

Qy 996 GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG 1048  
 | | | | | | | | | | | | | | | | | | | | | |

Db 1079 GGGCTCATGAACTCCTACTTTTCATTCAGAGAAAAGTGAGGGGCTTGTGAAACAG 1132

RESULT 7

AAD24958

ID AAD24958 standard; cDNA; 1542 BP.

XX

AC AAD24958;

XX

DT 12-MAR-2002 (first entry)

XX

DE Human G-protein coupled receptor-3 (GCREC-3) cDNA.

XX

KW Human; G-protein coupled receptor-3; GCREC-3; therapy; cancer; stroke;  
 KW cell proliferative disorder; neurological; epilepsy; Parkinson's disease;  
 KW Alzheimer's disease; inflammation; thyroiditis; haemolytic anaemia; AIDS;  
 KW Acquired Immune Deficiency Syndrome; dementia; nootropic; cholelithiasis;  
 KW multiple sclerosis; atherosclerosis; angina pectoris; gastroenteritis;  
 KW diabetes; ulcer; viral infection; immunosuppressive; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 63..1202

FT /\*tag= a

FT /product= "Human GCREC-3 protein"

XX

PN WO200198351-A2.

XX

PD 27-DEC-2001.

XX

PF 15-JUN-2001; 2001WO-US19275.

XX

PR 16-JUN-2000; 2000US-212483P.

PR 22-JUN-2000; 2000US-213954P.

PR 29-JUN-2000; 2000US-215209P.

PR 07-JUL-2000; 2000US-216595P.

PR 14-JUL-2000; 2000US-218936P.

PR 19-JUL-2000; 2000US-219154P.

PR 21-JUL-2000; 2000US-220141P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Lal P, Baughn MR, Hafalia AJA, Nguyen DB, Gandhi AR, Kallick DA;  
 PI Griffin JA, Yue H, Khan FA, Patterson C, Lu DAM, Tribouley CM;  
 PI Lu Y, Walia NK, Graul R, Yao MG, Yang J, Ramkumar J, Au-Young J;  
 PI Elliott VS, Hernandez R, Walsh RT, Borowsky ML, Thornton M, He A;

XX

DR WPI; 2002-075627/10.

DR P-PSDB; AAE15633.

XX

PT Isolated human G-protein coupled receptor polypeptides and the use of  
 PT these sequences in the diagnosis, treatment and prevention of diseases





Qy	519	CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCTGAAC	578
Db	685	CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGAAGTTCTGGAGACCCCAACT	744
Qy	579	ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA	638
Db	745	ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA	804
Qy	639	TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA	698
Db	805	TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA	864
Qy	699	CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA	758
Db	865	CTGCTCTGCCCTTGAAAAGCCTCTCAACTGGTCATCATGGCAGTGGTAATCTTCTCTG	924
Qy	759	TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT	818
Db	925	TGCTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT	984
Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	985	GGAAGCAGTATCAGTGCCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTT	1044
Qy	876	TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACA	935
Db	1045	TGGCCTTTCTGAACAGTGTCAACCCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA	1104
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	1105	GGGACATGCTGATGAATCAACTGAGACACAACCTCAAATCCCTTACATCCTTTAGCAGAT	1164
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG	1048
Db	1165	GGGCTCATGAACCTCTACTTTTCATTTCAGAGAAAAGTGAGGGGCTTGTGAAACAG	1218

RESULT 8

ABS57291

ID ABS57291 standard; cDNA; 1338 BP.

XX

AC ABS57291;

XX

DT 30-JAN-2003 (first entry)

XX

DE cDNA encoding human adenosine receptor.

XX

KW Human; mammalian; adenosine receptor; G-protein coupled receptor;  
 KW GPCR; adenosine-mediated medical condition; vasodilation; hypotension;  
 KW reversal of tachycardia; chronic renal disease; thyroid disorder;  
 KW inflammation; asthma; hypertensive; antiarrhythmic; antiinflammatory;  
 KW antiasthmatic; gene; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

```

FT      CDS                1..1005
FT                                     /*tag= a
FT                                     /product= "Adenosine receptor"
XX
PN      US2002137887-A1.
XX
PD      26-SEP-2002.
XX
PF      17-JAN-2001; 2001US-0765034.
XX
PR      17-JAN-2001; 2001US-0765034.
XX
PA      (HEDR/) HEDRICK J A.
PA      (LACH/) LACHOWICZ J E.
PA      (WANG/) WANG W.
PA      (GUST/) GUSTAFSON E L.
XX
PI      Hedrick JA, Lachowicz JE, Wang W, Gustafson EL;
XX
DR      WPI; 2003-074992/07.
DR      P-PSDB; ABG72131.
XX
PT      Novel isolated mammalian adenosine receptor polypeptide useful for
PT      identifying an agonist or antagonist of the receptor for treating
PT      vasodilation, hypotension, chronic renal diseases, thyroid disorders
PT      and inflammation -
XX
PS      Example 1; Page 14-16; 19pp; English.
XX
CC      The present invention relates to the isolation of a mammalian
CC      (human) adenosine receptor, and the polynucleotide sequence
CC      encoding it. The cloned receptor resembles a member of the
CC      G-protein coupled receptor (GPCR) superfamily that contains
CC      7-transmembrane domains. The adenosine receptor is useful for
CC      identifying agonists and antagonists of the receptor, which may be
CC      useful for treating an adenosine-mediated medical condition. The
CC      adenosine receptor polypeptide sequence is also useful as an
CC      antigen to elicit antibody production in an immunologically
CC      competent host. An antibody which binds specifically to the
CC      adenosine receptor is useful for treating medical conditions caused
CC      or mediated by adenosine such as vasodilation, hypotension, reversal
CC      of tachycardia, chronic renal diseases, thyroid disorders and
CC      inflammation (e.g. asthma). The antibody can also be used to purify
CC      the adenosine receptor, or as a basis for immunoassays of the receptor.
CC      The polynucleotide sequence encoding the adenosine receptor is useful
CC      for producing vectors and host cells containing the vectors. It is
CC      also useful for measuring expression of a mammalian adenosine
CC      receptor gene in a biological sample. The present sequence encodes
CC      human adenosine receptor.
XX
SQ      Sequence 1338 BP; 370 A; 288 C; 265 G; 415 T; 0 other;

Query Match          38.3%; Score 590.8; DB 25; Length 1338;
Best Local Similarity 75.2%; Pred. No. 9.2e-140;
Matches 763; Conservative 0; Mismatches 247; Indels 4; Gaps 2;

Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98

```

Db	8	GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA	67
Qy	99	AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA	158
Db	68	AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGAAAATACCA	127
Qy	159	CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT	218
Db	128	TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT	187
Qy	219	TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT	278
Db	188	TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT	247
Qy	279	ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC	338
Db	248	ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC	307
Qy	339	ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC	398
Db	308	ATGCCAACCTCTATACCAGCATCTCTTCTCACTTTTATCAGCATAGATCGATACTTGA	367
Qy	399	TCATGAAGTACCCTTTCCGAGAACACTTCTACAAAAGAAGGAATTTGCCATTTTAATCT	458
Db	368	TAATTAAGTATCCTTTCCGAGAACACTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT	427
Qy	459	CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT	518
Db	428	CCTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC	487
Qy	519	CTGTCCCAAAAGAAGAGGGCAGTAAC'TGCATCGACTATGCAAGTTCTGGAAACCCCTGAAC	578
Db	488	CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCCAACT	547
Qy	579	ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA	638
Db	548	ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA	607
Qy	639	TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA	698
Db	608	TGTGTTTCTTTTATTACAAGATTGCTCTCTTCCTAAAGCAGAGGAATAGGCAGGTTGCTA	667
Qy	699	CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTA	758
Db	668	CTGCTCTGCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG	727
Qy	759	TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT	818
Db	728	TGCCTTTTACACCCTATCACGTATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT	787
Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	788	GGAAGCAGTATCAGTGCACCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTT	847
Qy	876	TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACA	935

Db	848	TGGCCTTTCTGAACAGTGTCAACAACCCTGTCTTCTATTTTCCTTTGGGAGATCACTTCA	907
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	908	GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT	967
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG	1048
Db	968	GGGCTCATGAACTCCTACTTTTCATTGAGAGAAAAGTGAGGGGCTTGTGAAACAG	1021



Db	1292	CTGCTCTGCCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG	1351
Qy	759	TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT	818
Db	1352	TGCTTTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT	1411
Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	1412	GGAAGCAGTATCAGTGCACCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGGCTT	1471
Qy	876	TGGCCTTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACA	935
Db	1472	TGGGCTTTTCTGAACAGTGTCAATCAACCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA	1531
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	1532	GGGACATGCTGATGAATCAACTGAGACACAACCTCAAATCCCTTACATCCTTTAGCAGAT	1591
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG	1048
Db	1592	GGGCTCATGAACCTCTACTTTTCATTGAGAGAAAAGTGAGGGGCTTGTGAAACAG	1645

```

FT          /*tag= e
FT          /note= "binding site for primer used to amplify
FT              cDNA for bacterial expression"
FT primer_bind 1085..1096
FT          /*tag= f
FT          /note= "binding site for primer used to amplify
FT              cDNA for baculovirus expression"
XX
PN WO9724929-A1.
XX
PD 17-JUL-1997.
XX
PF 11-JAN-1996; 96WO-US00392.
XX
PR 11-JAN-1996; 96WO-US00392.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Li Y;
XX
DR WPI; 1997-372505/34.
DR P-PSDB; AAW22732.
XX
PT Isolated human ATP receptor - agonists and antagonists of which are
PT useful in treatment of, e.g. asthma, hypertension, arterial
PT thrombosis and psychotic and neurological disorders
XX
PS Claim 7; Fig 1A-C; 53pp; English.
XX
CC A cDNA clone (AAT75146) codes for human ATP receptor (AAW22732), a
CC polypeptide structurally related to the G protein-coupled receptor
CC family. It was discovered in a human placenta cDNA library.
CC cDNA encoding the mature receptor, deposited as ATCC 97333, can
CC be expressed in bacterial (e.g. E. coli), mammalian (e.g. COS) or
CC insect (e.g. Sf9) host cells and used to screen for agonists and
CC antagonists useful in the treatment of a variety of disorders.
CC It can also be used to identify a mutation in an ATP receptor gene
CC and thus to diagnose diseases, or susceptibility to diseases,
CC related to ATP receptor underexpression.
XX
SQ Sequence 1428 BP; 394 A; 308 C; 290 G; 435 T; 1 other;

Query Match          38.1%; Score 587.2; DB 18; Length 1428;
Best Local Similarity 75.0%; Pred. No. 7.8e-139;
Matches 760; Conservative 1; Mismatches 249; Indels 4; Gaps 2;

Qy          39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db          99 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 158

Qy          99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTGGACTGCTTGGGAATGTCA 158
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db         159 AGTACTACCTTTCCATTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 218

Qy         159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db         219 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 278

```



Qy	219	TTAACCTTTCCATCTCTGACTTTGCTTTCCCTGTGCACCTTCCCATCCTGATAAAGAGTT	278
Db	279	TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCTCCCCATGCTGATAAGGAGTT	338
Qy	279	ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC	338
Db	339	ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC	398
Qy	339	ACACCAACCTCTACACCAGCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGC	398
Db	399	ATGCCAACCTCTATACCAGCATTCTCTTCTCACTTTTATCAGCATAGATCGATAC TTGA	458
Qy	399	TCATGAAGTACCCTTTCCGAGAACACTTCTACAAAAGAAGGAATTTGCCATTTTAAATCT	458
Db	459	TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTGTGCTATTTTAAATCT	518
Qy	459	CGCTGGCTGTCTGGGCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT	518
Db	519	CCTTGGCCATGTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC	578
Qy	519	CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAAC	578
Db	579	CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCACT	638
Qy	579	ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTGTGA	638
Db	639	ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCCTCTTTTGTGA	698
Qy	639	TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA	698
Db	699	TGTGTTTCTTTTATTACAAGATTGCCTCCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA	758
Qy	699	CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTA	758
Db	759	CTGCCTCGCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG	818
Qy	759	TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT	818
		:	
Db	819	TGCTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT	878
Qy	819	G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC	875
Db	879	GGAAGCAGTATCAGTGCACCTCAGGTCGTCATCAACTCCTTTTACATTGTGACACGGCCTG	938
Qy	876	TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCCTCATGGGAGACCATTACA	935
Db	939	TGGCCTTTCTGAACAGTGTCAATCAACCCTGTCTTCTATTTTCTTGTGGGAGATCACTTCA	998
Qy	936	GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT	995
Db	999	GGGACATGCTGATGAATCAACTGAGACACAACCTTCAAATCCCTTACATCCTTTAGCAGAT	1058
Qy	996	GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG	1048
Db	1059	GGGCTCATGAACTCCTACTTTTCATTTCAGAGAAAAGTGAGGGGCTTGTGAAACAG	1112

RESULT 11

AAC81122

ID AAC81122 standard; cDNA; 1385 BP.

XX

AC AAC81122;

XX

DT 14-FEB-2001 (first entry)

XX

DE Human secreted protein gene 37 SEQ ID NO:47.

XX

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
KW fungicide; ophthalmological; vulnerary; gene therapy; autoimmune disease;  
KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;  
KW cerebrovascular disorder; nervous system disorder; infection; skin aging;  
KW ocular disorder; wound healing; food additive; preservative; ss.

XX

OS Homo sapiens.

XX

PN WO200061628-A1.

XX

PD 19-OCT-2000.

XX

PF 06-APR-2000; 2000WO-US09070.

XX

PR 09-APR-1999; 99US-0128695.

PR 14-JAN-2000; 2000US-0176052.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI; 2000-619228/59.

DR P-PSDB; AAB45344.

XX

PT New nucleic acid molecules encoding 49 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -

XX

PS Claim 1; Page 412; 454pp; English.

XX

CC The polynucleotide sequences given in AAC81086 to AAC81134 encode the  
CC human secreted proteins given in AAB45308 to AAB45356. AAB45357 to  
CC AAB45384 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;  
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
CC and vulnerary. The polynucleotides and polypeptides can be used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Disorders which are diagnosed or treated include

CC autoimmune diseases, hyperproliferative disorders, cardiovascular  
CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound healing and  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
CC components. AAC81077 to AAC81085 and AAB45307 represent sequences used in  
CC the exemplification of the present invention.

XX

SQ Sequence 1385 BP; 385 A; 296 C; 275 G; 429 T; 0 other;

Query Match 37.6%; Score 580.4; DB 21; Length 1385;  
Best Local Similarity 75.2%; Pred. No. 4.1e-137;  
Matches 763; Conservative 0; Mismatches 246; Indels 5; Gaps 3;

```
Qy      39 GCAGAATGGCACAGAATTTATCTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATA 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      56 GGATCATGGCATGGAATGCAACTTGCAAAACTGGCTGGCAGCAGAGGCTGCCCTGGAAA 115

Qy      99 AGTACTACCTCTCTGCATTTTATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCA 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     116 AGTACTACCTTTCCATTTTTTATGGGATTGAGTTCGTTGTGGGAGTCCTTGGAAATACCA 175

Qy     159 CTGTGGTGTTCGGCTACCTCTTCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTT 218
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     176 TTGTTGTTTACGGCTACATCTTCTCTCTGAAGAACTGGAACAGCAGTAATATTTATCTCT 235

Qy     219 TTAACCTTTCCATCTCTGACTTTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTT 278
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     236 TTAACCTCTCTGTCTCTGACTTAGCTTTTCTGTGCACCCTCCCCATGCTGATAAGGAGTT 295

Qy     279 ATGCCAATGATAAGGGGACCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTC 338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     296 ATGCCAATGGAACTGGATATATGGAGACGTGCTCTGCATAAGCAACCGATATGTGCTTC 355

Qy     339 ACACCAACCTCTACACCAGCATCCTCTTCTCACTTTTCATTAGCATGGACCGATATCTGC 398
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     356 ATGCCAACCTCTATACCAGCATCTCTTCTCACTTTTATCAGCATAGATCGATACTTGA 415

Qy     399 TCATGAAGTACCCTTTCCGAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCT 458
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     416 TAATTAAGTATCCTTTCCGAGAACACCTTCTGCAAAAGAAAGAGTTTGCTATTTTAATCT 475

Qy     459 CGCTGGCTGTCTGGGCCCTTAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATT 518
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     476 CTTGGCCATTTGGGTTTTAGTAACCTTAGAGTTACTACCCATACTTCCCCTTATAAATC 535

Qy     519 CTGTCCCAAAAGAAGAGGGCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCCTGAAC 578
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     536 CTGTTATAACTGACAATGGCACCACCTGTAATGATTTTGCAAGTTCTGGAGACCCCAACT 595

Qy     579 ACAATCTCATTTACAGCCTCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGA 638
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 596 ACAACCTCATTTACAGCATGTGTCTAACACTGTTGGGGTTCCTTATTCTCTTTTTGTGA 655

Qy 639 TGTGCTTCTTCTACTACAAGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAA 698  
 |||| |||| | ||||| | |||| |||| |||| |||| |||

Db 656 TGTGTTTCTTTTATTACAAGATTGCTCTCTTCTTAAAGCAGAGGAATAGGCAGGTTGCTA 715

Qy 699 CTGCCCTGCCACTGGACAAACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTA 758  
 |||| |||| | || || || | |||| || |||| || || |||||

Db 716 CTGCTCTGCCCCTTGAAAAGCCTCTCAACTTGGTCATCATGGCAGTGGTAATCTTCTCTG 775

Qy 759 TACTCTTCACACCCTATCATATCATGCGCAATTTGAGGATCGCCTCACGCCTGGATAGTT 818  
 | || || ||||| ||||| || ||||| ||||| ||||| ||||

Db 776 TGCTTTTACACCCTATCACGTCATGCGGAATGTGAGGATCGCTTCACGCCTGGGGAGTT 835

Qy 819 G---GCCACAAGGATGTACACAGAAGGCCATCAAATCTATATACACACTGACACGGCCTC 875  
 | || | || || || | |||| || | |||| |||| ||||

Db 836 GGAAGCAGTATCAGTGCCTCAGGTCGTCACTCAACTCCTTTTACATTGTGACAC-GCCTT 894

Qy 876 TGGCCTTTCTGAACAGTGCCATCAATCCCATCTTCTACTTCCTCATGGGAGACCATTACA 935  
 ||||| ||||| ||||| || ||||| || || ||||| || || ||

Db 895 TGGCCTTTCTGAACAGTGCATCAACCCTGTCTTCTATTTTCTTTTGGGAGATCACTTCA 954

Qy 936 GAGAGATGCTGATTAGTAAGTTCAGACAATACTTCAAGTCCCTTACATCCTTCAGGACAT 995  
 | || ||||| || | | |||| ||||| ||||| ||||| || || ||

Db 955 GGGACATGCTGATGAATCAACTGAGACACAACCTCAAATCCCTTACATCCTTTAGCAGAT 1014

Qy 996 GAGCTGCTGGATGCAGGTCTTCACTCAGCCAAAA-TGAGACACTTGATAAACAG 1048  
 | ||| || | | |||| |||| |||| |||| |||| |||||

Db 1015 GGGCTCATGAACTCCTACTTTTCATTTCAGAGAAAAGTGAGGGGCTTGTGAAACAG 1068

# RESULT 12

AAL43942

ID AAL43942 standard; DNA; 1011 BP.

XX

AC AAL43942;

XX

DT 27-SEP-2002 (first entry)

XX

DE Human G protein-coupled receptor coding sequence.

XX

KW Human; gene therapy; G protein-coupled receptor; drug development;  
 KW central nervous system disease; endocrine disease; metabolic disease;  
 KW cancer; respiratory disease; digestive disease; immune disease;  
 KW inflammation; infection; circulatory disease; gene; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1..1011

FT /\*tag= a

FT /partial

FT /product= "Human G-protein coupled receptor"

FT /note= "No stop codon is given"

XX

PN WO200257441-A1.

XX

```

PD      25-JUL-2002.
XX
XX
PF      17-JAN-2002; 2002WO-JP00270.
XX
PR      18-JAN-2001; 2001JP-0010714.
PR      30-MAR-2001; 2001JP-0102484.
XX
PA      (TAKE ) TAKEDA CHEM IND LTD.
XX
PI      Miwa M, Ito T, Shintani Y, Miyajima N;
XX
DR      WPI; 2002-566800/60.
DR      P-PSDB; AAO15399.
XX
PT      Human kidney-originated G protein-coupled receptor protein TGR30 and
PT      encoded DNA, for developing drugs to treat central nervous diseases,
PT      endocrine diseases, metabolic diseases and cancer, including gene
PT      therapy -
XX
PS      Claim 6; Page 90-91; 98pp; Japanese.
XX
CC      The invention comprises the amino acid and coding sequence of a human G
CC      protein-coupled receptor. The DNA and protein sequences of the invention
CC      are useful for developing drugs to prevent or treat (gene therapy):
CC      central nervous system diseases; endocrine diseases; metabolic diseases;
CC      cancer; respiratory diseases; digestive diseases; immune diseases;
CC      inflammations; infections; and circulatory diseases. The present DNA
CC      sequence encodes the human G protein-coupled receptor of the invention.
XX
SQ      Sequence 1011 BP; 257 A; 263 C; 188 G; 303 T; 0 other;

Query Match          8.2%; Score 126.6; DB 24; Length 1011;
Best Local Similarity 49.9%; Pred. No. 6.2e-22;
Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      ||| || ||||| | | | | | ||| ||| |||
Db      59 CTTTTGGAAATTGCACTGATGAAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTT 118

Qy      120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCAGTGTGGTGTTCGGCTACCTCT 179
      ||| || ||| | | ||| | | ||| ||| ||| |||
Db      119 ATGGCATTATCTTCCTCGTGGGATTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178

Qy      180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTTAACCTTTCCATCTCTGACT 239
      || |||| |||| ||||| || | | | |||| || | | |
Db      179 TCAAATGAGACCTTGGAAGAGCAGCACCATCATATGCTGAACCTGGCCTGCACAGATC 238

Qy      240 TTGCTTTTCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296
      | | | || || || || ||||| | ||||| | || | ||
Db      239 TGCTGTATCTGACCAGCCTCCCCTTCCTGATTCACTACTATGCCAGTGGCGAAACTGGA 298

Qy      297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      || ||||| | | ||| | ||| | || ||||| || | ||
Db      299 TCTTTGGAGATTTTATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCA 358

Qy      357 GCATCCTCTTCTCACTTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      ||||| ||||| || | |||| ||| || | ||| | |||| |

```

```

Db      359 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTACCCAATGA 418
Qy      417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      419 GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478
Qy      477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGG 536
      |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      479 TTTCACTGGTAGCTGTCAATCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538
Qy      537 GCAGTAACTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596
      |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      539 GATCAGCCTGTCTCGACCTCACCAGTTCGG-----ATGAACCAATACTATTAAGTGGT 592
Qy      597 TCTGCCTGACTTTGTGGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656
      |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      593 ACAACCTGATTTTGACTGCAACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652
Qy      657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACCTGACAGCTGCCTTAAGCAGA 712
Qy      717 AACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776
      |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      713 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTACGTATGTTTTTACCCTTCC 772
Qy      777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811
      |   |   |   |   |   |   |   |   |   |   |   |   |
Db      773 ATATCTTGAGGGTCATTCGGATCGAATCTCGCCTG 807

```

# RESULT 13

AAS07948

ID AAS07948 standard; cDNA; 1014 BP.

XX

AC AAS07948;

XX

DT 23-OCT-2001 (first entry)

XX

DE Human cDNA encoding G-protein coupled receptor, hRUP21.

XX

KW Human; G-protein coupled receptor; GPCR; hRUP21; agonist;

KW inverse agonist; lung cancer; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1..1014

FT /\*tag= a

FT /product= "hRUP21"

XX

PN WO200136471-A2.

XX

PD 25-MAY-2001.

XX

PF 16-NOV-2000; 2000WO-US31509.

XX

PR 17-NOV-1999; 99US-0166088.  
PR 17-NOV-1999; 99US-0166099.  
PR 17-NOV-1999; 99US-0166369.  
PR 23-DEC-1999; 99US-0171900.  
PR 23-DEC-1999; 99US-0171901.  
PR 23-DEC-1999; 99US-0171902.  
PR 11-FEB-2000; 2000US-0181749.  
PR 14-MAR-2000; 2000US-0189258.  
PR 14-MAR-2000; 2000US-0189259.  
PR 10-APR-2000; 2000US-0195898.  
PR 10-APR-2000; 2000US-0195899.  
PR 10-APR-2000; 2000US-0196078.  
PR 28-APR-2000; 2000US-0200419.  
PR 12-MAY-2000; 2000US-0203630.  
PR 12-JUN-2000; 2000US-0210741.  
PR 12-JUN-2000; 2000US-0210982.  
PR 21-AUG-2000; 2000US-0226760.  
PR 26-SEP-2000; 2000US-0235418.  
PR 26-SEP-2000; 2000US-0235779.  
PR 20-OCT-2000; 2000US-0242332.  
PR 20-OCT-2000; 2000US-0242343.

XX

PA (AREN-) ARENA PHARM INC.

XX

PI Chen R, Dang HT, Lowitz KP;

XX

DR WPI; 2001-355616/37.

DR P-PSDB; AAU04375.

XX

PT Endogenous and non-endogenous versions of human G-protein coupled  
PT receptors for direct identification of candidate compounds as agonists,  
PT inverse agonists or partial agonists for use as therapeutic agents -

XX

PS Claim 55; Page 113-114; 159pp; English.

XX

CC The sequence encodes a human G-protein coupled receptor (GPCR),  
CC hRUP21 The endogenous and non-endogenous, constitutively activated  
CC versions of human G-protein coupled receptors (GPCR), are useful for  
CC direct identification of candidate compounds as receptor agonists,  
CC inverse agonists or partial agonists having applicability as therapeutic  
CC agents for treating diseases related to GPCR, e.g. lung cancer.  
CC Non-endogenous version of human GPCRs are also utilized in research  
CC settings and in vitro and in vivo system, incorporating GPCRs can be  
CC utilised to elucidate and understand the roles these receptors  
CC play in the human condition, both normal and diseased.

XX

SQ Sequence 1014 BP; 258 A; 263 C; 189 G; 304 T; 0 other;

Query Match 8.2%; Score 126.6; DB 22; Length 1014;

Best Local Similarity 49.9%; Pred. No. 6.2e-22;

Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

Qy 60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119  
||| || ||||| | | | | | | | | | | | | | |  
Db 59 CTTTGGAAATTGCACTGATGAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTT 118  
  
Qy 120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCT 179

```

      |||  ||   |||  |  | |||  |  | |||  ||   |||  |  |  |||  |  |
Db      119 ATGGCATTATCTTCCTCGTGGGATTTCAGGCAATGCAGTAGTGATATCCACTTACATTT 178

Qy      180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTTTCCATCTCTGACT 239
      ||   |||  |||  |||  |||  |||  ||   ||   ||   |||  ||   ||   ||
Db      179 TCAAAATGAGACCTTGGAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATC 238

Qy      240 TTGCTTTCCTGTGACCCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296
      |   |   |||  ||   |||  |||  |||  |||  ||   |||  ||   |||  |||
Db      239 TGCTGTATCTGACCAGCCTCCCCTTCCTGATTCACTACTATGCCAGTGGCGAAAACCTGGA 298

Qy      297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      ||   |||  |||  |   |||  ||   |||  ||   ||   |||  ||   ||   ||
Db      299 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCA 358

Qy      357 GCATCCTCTTCCTCACCTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      |||  |||  |||  |||  ||   ||   |||  ||   ||   ||   ||   ||   ||
Db      359 GCATCCTCTTCCTCACCTGTTTCAGCATCTCCGCTACTGTGTGATCATTCACCCAATGA 418

Qy      417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476
      |   |   |   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      419 GCTGCTTTTCCATTACAAAACTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478

Qy      477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGG 536
      |   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      479 TTTCACCTGGTAGCTGTCAATCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538

Qy      537 GCAGTAACCTGCATCGACTATGCAAGTTCTGGAAAACCTGAACACAATCTCATTTACAGCC 596
      |   ||   |||  |||  ||   |||  ||   |||  |||  |||  ||   ||   ||
Db      539 GATCAGCCTGTCTCGACCTCACCAAGTTCGG-----ATGAACTCAATACTATTAAGTGGT 592

Qy      597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656
      |   |||  |||  ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      593 ACAACCTGATTTTGACTGCAACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652

Qy      657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACTGACAGCTGCCTTAAGCAGA 712

Qy      717 AACCCCAACGCCTGGTGGTCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776
      ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      713 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTACGTATGTTTTTTACCCTTCC 772

Qy      777 ATATCATGCGCAATTTGAGGATCGCCTCAGCCTG 811
      |||  ||   ||   ||   |||  ||   |||  |||  |||  |||
Db      773 ATATCTTGAGGGTCATTCGGATCGAATCTCGCCTG 807

```

RESULT 14

ABZ42876

ID ABZ42876 standard; DNA; 1014 BP.

XX

AC ABZ42876;

XX

DT 06-MAR-2003 (first entry)

XX

DE Human GPCR polynucleotide SEQ ID NO 13.



XX  
 KW Human; GPCR; G protein coupled receptor; signal transduction; olfactory;  
 KW drug development; gustatory; taste; fragrance; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200216548-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 30-JUL-2001; 2001WO-IB01446.  
 XX  
 PR 04-AUG-2000; 2000JP-0237818.  
 PR 13-FEB-2001; 2001JP-0034434.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Haga T, Takeda S, Mitaku S;  
 XX  
 DR WPI; 2002-304118/34.  
 DR P-PSDB; ABP95602.  
 XX  
 PT Database global search for G protein-coupled receptors, proteins and  
 PT encoded genes for studying in vivo signal transduction mechanism and  
 PT identifying targets for drug development -  
 XX  
 PS Claim 9; SEQ ID NO 13; 97pp + Sequence Listing; Japanese.  
 XX  
 CC The invention relates to a method for screening G protein-coupled  
 CC receptor (GPCR) genes (ABZ42870-ABZ43216) and/or GPCR proteins  
 CC (ABP95596-ABP95942) by extracting open-reading frames containing 6-8  
 CC transmembrane domains with 250-1000 amino acid residues to give a gene  
 CC homologous with a known GPCR gene. The receptor proteins and encoded  
 CC genes are useful for studying in vivo signal transduction mechanism and  
 CC identifying targets for drug development e.g. based on olfactory and  
 CC gustatory receptors in form of agonists and antagonists by screening  
 CC intrinsic and extrinsic ligands as bitter taste inhibitors, taste  
 CC enhancers and fragrance improvers.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1014 BP; 258 A; 263 C; 189 G; 304 T; 0 other;

Query Match 8.2%; Score 126.6; DB 24; Length 1014;  
 Best Local Similarity 49.9%; Pred. No. 6.2e-22;  
 Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

Qy 60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119  
 ||| || |||| | | | ||| ||| |||  
 Db 59 CTTTGGAAATTGCACTGATGAAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTT 118  
  
 Qy 120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACCTGTGGTGTTCGGCTACCTCT 179  
 ||| || ||| | | ||| | | || ||| ||| |||  
 Db 119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178  
  
 Qy 180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTAAACCTTCCATCTCTGACT 239

Db 179 TCAAAATGAGACCTTGGAAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATC 238  
 Qy 240 TTGCTTTCCTGTGCACCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296  
 Db 239 TGCTGTATCTGACCAGCCTCCCCTTCCTGATTCACTACTATGCCAGTGGCGAAAACCTGGA 298  
 Qy 297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356  
 Db 299 TCTTTGGAGATTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCA 358  
 Qy 357 GCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416  
 Db 359 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTACCCCAATGA 418  
 Qy 417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476  
 Db 419 GCTGCTTTTCCATTACAAAACTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478  
 Qy 477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGG 536  
 Db 479 TTTCACCTGGTAGCTGTCAATCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538  
 Qy 537 GCAGTAACCTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596  
 Db 539 GATCAGCCTGTCTCGACCTCACCAAGTTCGG-----ATGAACTCAATACTATTAAGTGGT 592  
 Qy 597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656  
 Db 593 ACAACCTGATTTTGACTGCAACTACTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652  
 Qy 657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716  
 Db 653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACTGACAGCTGCCTTAAGCAGA 712  
 Qy 717 AACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776  
 Db 713 AAGCACGAAGGCTAACCATTCTGCTACTCCTTGCAATTTACGTATGTTTTTACCCTTCC 772  
 Qy 777 ATATCATGCGCAATTTGAGGATCGCCTCACGCCTG 811  
 Db 773 ATATCTTGAGGGTCATTCGGATCGAATCTCGCCTG 807

# RESULT 15

ABN85630

ID ABN85630 standard; DNA; 1014 BP.

XX

AC ABN85630;

XX

DT 18-SEP-2002 (first entry)

XX

DE Human P2Y-like receptor variant encoding gene SEQ ID NO 3.

XX

KW Human; Py2-like receptor; HIPHUM 0000037; immunity; inflammation;

KW cancer; Crohn's disease; irritable bowel syndrome; rheumatoid arthritis;

KW immunomodulator; anti-inflammatory; cytostatic; antiasthmatic;

KW gastrointestinal; anti-ulcer; antirheumatic; antiarthritic; virucide;  
 KW antibacterial; immunosuppressive; dermatological; nephrotropic;  
 KW antiallergic; analgesic; receptor; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1014  
 FT /\*tag= a  
 FT /product= "P2Y-like receptor variant"  
 XX  
 PN GB2369364-A.  
 XX  
 PD 29-MAY-2002.  
 XX  
 PF 31-AUG-2001; 2001GB-0021215.  
 XX  
 PR 01-SEP-2000; 2000GB-0021524.  
 PR 06-SEP-2000; 2000GB-0021894.  
 PR 25-SEP-2000; 2000GB-0023444.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Foord SM, Ignar DM;  
 XX  
 DR WPI; 2002-511268/55.  
 DR P-PSDB; ABB83819.  
 XX  
 PT An isolated P2Y-like receptor polypeptide (HIPHUM 0000037) which can be  
 PT used for the identification of agonists and antagonists which may be  
 PT used to treat an immune or inflammatory disease -  
 XX  
 PS Claim 5; Page 28-29; 35pp; English.  
 XX  
 CC The invention relates to an isolated P2Y-like receptor polypeptide  
 CC (ABB83818-ABB83819) which is also referred to in the specification as  
 CC HIPHUM 0000037. An effective amount of a substance (agonist or  
 CC antagonist) which modulates P2Y receptor activity is useful to treat a  
 CC subject having a disorder that is responsive to P2Y-like receptor  
 CC modulation. The disorder is a disease of immunity or inflammation. The  
 CC substance may also be used to manufacture a medicine for the treatment or  
 CC prophylaxis of a disorder that is responsive to stimulation or modulation  
 CC of P2Y-like receptor activity. Disorders which may be treated include  
 CC colon cancers, asthma, COPD, Crohn's disease, irritable bowel syndrome,  
 CC gastroenteritis and colitis, inflammatory bowel syndrome, ulcerative  
 CC colitis, rheumatoid arthritis, viral diseases, bacterial infections,  
 CC autoimmune diseases, dermatitis, glomerulonephritis allergies, allergic  
 CC rhinitis, inflammatory pain and general inflammation such as tendonitis,  
 CC polymyositis or prostatitis. The invention provides alternative  
 CC substances for the treatment of immunological and inflammatory diseases.  
 CC The present sequence is that the P2Y-like receptor variant encoding gene  
 CC of the invention.  
 XX  
 SQ Sequence 1014 BP; 258 A; 263 C; 189 G; 304 T; 0 other;

Query Match 8.2%; Score 126.6; DB 24; Length 1014;  
 Best Local Similarity 49.9%; Pred. No. 6.2e-22;

Matches 377; Conservative 0; Mismatches 369; Indels 9; Gaps 2;

```

Qy      60 CTTGTGAGAATTGGTTGGCAACAGAGGCTATCTTGAATAAGTACTACCTCTCTGCATTTT 119
      ||| || ||||| | | | | | | | | | | | | | |
Db      59 CTTTGGAAATTGCACTGATGAAAACATCCCACTCAAGATGCACTACCTCCCTGTTATTT 118

Qy     120 ATGCAATCGAGTTCATTTTTGGACTGCTTGGGAATGTCACTGTGGTGTTCGGCTACCTCT 179
      ||| || ||| | | ||| | | |||| | | | | | | | |
Db     119 ATGGCATTATCTTCCTCGTGGGATTTCCAGGCAATGCAGTAGTGATATCCACTTACATTT 178

Qy     180 TCTGCATGAAGAACTGGAACAGCAGCAATGTCTATCTTTTTAACCTTTCCATCTCTGACT 239
      || |||| | |||| | |||| | | | | | | | | | |
Db     179 TCAAAATGAGACCTTGGAAGAGCAGCACCATCATTATGCTGAACCTGGCCTGCACAGATC 238

Qy     240 TTGCTTTCCTGTGCACCCCTTCCCATCCTGATAAAGAGTTATGCCAAT---GATAAGGGGA 296
      | | | | | | | | | | | | | | | | | | | | |
Db     239 TGCTGTATCTGACCAGCCTCCCCTTCTGATTCACTACTATGCCAGTGGCGAAAACCTGGA 298

Qy     297 CCTATGGAGATGTTCTCTGTATAAGCAACCGATATGTGCTTCACACCAACCTCTACACCA 356
      || ||||| | | |||| | ||| | | | | | | | | |
Db     299 TCTTTGGAGATTTTCATGTGTAAGTTTATCCGCTTCAGCTTCCATTTCAACCTGTATAGCA 358

Qy     357 GCATCCTCTTCCTCACTTTCATTAGCATGGACCGATATCTGCTCATGAAGTACCCTTTCC 416
      ||||| ||||| | | |||| | ||| | | | | | | |
Db     359 GCATCCTCTTCCTCACCTGTTTCAGCATCTTCCGCTACTGTGTGATCATTCACCCAATGA 418

Qy     417 GAGAACACTTTCTACAAAAGAAGGAATTTGCCATTTTAATCTCGCTGGCTGTCTGGGCCT 476
      | | | | | | | | | | | | | | | | | | | | |
Db     419 GCTGCTTTTCCATTACAAAACCTCGATGTGCAGTTGTAGCCTGTGCTGTGGTGTGGATCA 478

Qy     477 TAGTGACCTTAGAAGTTCTACCCATGCTCACTTTCATCAATTCTGTCCCAAAGAAGAGG 536
      | | | | | | | | | | | | | | | | | | | |
Db     479 TTTCACTGGTAGCTGTCACTCCGATGACCTTCTTGATCACATCAACCAACAGGACCAACA 538

Qy     537 GCAGTAACCTGCATCGACTATGCAAGTTCTGGAAACCCTGAACACAATCTCATTTACAGCC 596
      | | | | | | | | | | | | | | | | | | | |
Db     539 GATCAGCCTGTCTCGACCTCACCAGTTTCGG-----ATGAACCTCAATACTATTAAGTGGT 592

Qy     597 TCTGCCTGACTTTGTTGGGCTTCCTAATTCCTCTCTCTGTGATGTGCTTCTTCTACTACA 656
      | |||| | || | | | | | | | | | | | | | |
Db     593 ACAACCTGATTTTGACTGCAACTACTTTCTGCCTCCCCTTGGTGATAGTGACACTTTGCT 652

Qy     657 AGATGGTAGTCTTCTTAAAGAGGAGGAGCCAGCAGCAAGCAACTGCCCTGCCACTGGACA 716
      || | | | | | | | | | | | | | | | | | |
Db     653 ATACCACGATTATCCACACTCTGACCCATGGACTGCAAACTGACAGCTGCCTTAAGCAGA 712

Qy     717 AACCCCAACGCCTGGTGGTCCTGGCGGTTGTGATCTTCTCTATACTCTTCACACCCTATC 776
      || | | | | | | | | | | | | | | | | | |
Db     713 AAGCACGAAGGCTAACCATTTCTGCTACTCCTTGCAATTTACGTATGTTTTTACCCTTCC 772

Qy     777 ATATCATGCGCAATTTGAGGATCGCCTCAGCCTG 811
      |||| | | | | | | | | | | | | |
Db     773 ATATCTTGAGGGTCATTCGGATCGAATCTCGCCTG 807

```

Search completed: December 14, 2003, 13:22:08  
Job time : 453 secs